Quantitative Structure Activity Studies As A Tool For Molecular Designing Of More Potent Drugs

A

Thesis

Submitted to the

University of Allahabad

For the Degree of

Doctor of Philosophy

Submitted by

Soumya Srivastava

Supervisor

Professor K. P. Tiwari

Department of Chemistry
University of Allahabad
Allahabad (India)

2002

Dr. K. P. Tiwari
F.N.A.Sc.
Professor & Head

Department of Chemistry
University of Allahabad
Allahabad-211002, INDIA
Phone : 609936

Certificate

This is to certify that ***Ms. Soumya Srivastava*** *has worked under my supervision and has submitted a thesis entitled* ***"Quantitative Structure Activity Studies As A Tool For Molecular Designing Of More Potent Drugs"*** *for D.Phil. degree in the University of Allahabad, Allahabad. The work is original and has not been submitted for any degree elsewhere.*

K P Tiwari

Place : Allahabad
Date : 12.12.02

(Professor K. P. Tiwari)

Residence : 17/7A, Kasturba Gandhi Marg, Allahabad - 211002 • Phone : (0532) 608220

PREFACE

The present work entitled *"Quantitative structure Activity Studies As A Tool For Molecular Designing of More Potent Drugs"* deals with 2D and 3D QSAR studies of some steroidal and non steroidal estrogen receptor. The thesis has been divided into four chapters.

The *Chapter-1* has been divided into three sections. The *Section-A* gives an insight on the various aspects of drug designing and computer aided molecular modelling. The *Section-B* describes about QSAR (2D) methodology and parametrization. 3D-QSAR approaches have also been incorporated in this section. The *Section-C* includes the regression analysis method employed in this work for obtaining the QSAR along with a detailed description of the various descriptors used in the present thesis.

The *Chapter-2* comprised of an overview of the introductory idea about estrogen receptors and the various selective estrogen receptor modulators on which the present work has been carried out.

The *Chapter-3* has been divided into two sections. The *Section-A* is concerned with the results, discussion and conclusions part of SERMs placed on a common relative binding affinity (RBA) scale and the *Section-B* is concerned with the results, discussion and conclusions part of the SERMs whose binding affinity values are expressed in IC_{50} forms on which the present work has been carried out.

The *Chapter-4* comprises of two sections. The *Section-A* deals with introduction to 3D-QSAR using APEX-3D and CATALYST. The *Section-B* includes the results, discussion and conclusions of the estrogen receptors taken for 3D-QSAR study. This section also includes molecular modelling of some new potent estrogen receptors. In the last, references

and certificate of training programme done on QSAR and molecular modelling at CDRI, Lucknow are also included.

The investigations incorporated in the thesis were carried out in the *Chemistry laboratory of the Department of Chemistry, University of Allahabad.* The results of research work have not so for been submitted in part or in full for any degree or diploma of any Univeristy.

A summary of the entire work incorporated is being submitted separately along with the thesis as required by the ordinance of the University of Allahabad for the award of D.Phil. Degree.

Date: 12/12/02

(Ms. Soumya Srivastava)
Research Scholar
Department of Chemistry
University of Allahabad
Allahabad-211 002

ACKNOWLEDGEMENT

I bow in reverence to God whose benign benediction gave me the required zeal and confidence to complete this work.

Words are too scarce to place on record my heartiest gratitude and indebtedness to my mentor and esteemed supervisor ***Professor K. P. Tiwari, F.N.A.Sc., Ex-Head, Department of Chemistry University of Allahabad, Allahabad (U.P.)*** for his invaluable guidance, incessant encouragement and affectionate behaviour. His gentle demeanour and loving attitude helped me to overcome many hurdles and made it possible for me to bring this humble attempt to completion. My heartiest gratitude I owe to him is indeed beyond my expression.

I am particularly grateful to ***Dr. C. M. Gupta, Director, Central Drug Research Institute, Lucknow (U.P.)***, to have granted me permission to under go training at Medicinal Chemistry Division, CDRI, Lucknow. My vocabulary falls short to record my deep sense of gratitude to ***Dr. A. K. Saxena, Deputy Director, Medicinal Chemistry Division, Central Drug Research Institute (CDRI), Lucknow (U.P.)*** for providing me the opportunity to join training program and to learn 3D-QSAR and Molecular Modelling under his valued guidance. His continuous interest in my work will remain a source of inspiration for me.

It gives me immense pleasure in expressing my sincere gratitude and special thanks to ***Professor A. K. Srivastava, Deptt. of Chemistry Allahabad University, Allahabad*** for his suggestions, cooperation and encouragement throughout the present course of investigation.

I feel proud to express my profound gratitude to ***Professor Bharat Singh, Head, Deptt. of Chemistry, Allahabad University, Allahabad*** for his generous encouragement and for provision of all facilities essentially required during this period.

I feel it my humble and pious duty to mention the sagacious and brilliant guidance of my parents, ***Professor S. K. Srivastava, Department of Chemistry, Dr. H. S. G. University, Sagar (M.P.)*** and ***Dr. Mrs. S. D. Srivastava. Reader, Department of Chemistry, Dr. H.S.G. University, Sagar (M.P.)*** for providing me opportunity and new insights at every stage of the investigation. They have been the constant source of inspiration, encouragement, eugenicity and knowledge. Without their multifarious help, the present work would not have been possible for its completion.

My heartfelt thanks to the research group of Dr. A. K. Saxena and specially to Mr. Philip Prathipathi, SRF-CSIR, for their unfathomable help extended to me in various ways at every juncture of my work.

I must record my sincere respects and regards to my grandfather and grandmother for whom words become meaningless but for mere formality I have to express my deepest sense of gratitude for their affectionate patronage. I am deeply beholden to my sister ***Er. Sonal*** and brother ***Er. Sourabh*** who have given their full cooperation and help for the successful completion of my Ph.D. work.

Every success story begins with a dream. This work would have remained a dream, the idea of a notion, if it had not been for the guidance and cooperation I received. Last but not the least I convey my indebtedness to all those who have directly or indirectly contributed to the successful completion of my Ph.D. work.

Dated: 12/12/02

Soumya..

Ms. SOUMYA SRIVASTAVA
Research Scholar
Department of Chemistry
University of Allahabad
Allahabad (U.P.)
INDIA

Contents

CHAPTER-1

INTRODUCTION

SECTION-A

DRUG DESIGN

Drugs are of major importance to human health and nutrition on account of their action on biological and pathological process. The use of drugs has shown considerable worldwide increase in recent years and this tendency is likely to be increased in near future. In the earlier days purely randomized search procedures were involved in the discovery of new drugs. But the randomized search is no longer effective, as it is too time consuming, guarantees too title success and is too expensive. The chance of discovering a new agent has diminished to 1 in 10,000 and will decrease even further. Development costs has risen to more than 40 million dollars per new drug. This necessitated the development of a new logical and scientific approach in discovery of new drug, which is known as **"Drug Design"**[1].

Drug design is an integrated developing discipline which portends an era of tailored drugs. Tailoring of drug[2] means alteration of various physical and chemical properties of drug molecule through insertion of newer functional moieties or by the replacement of such groups already present by others for example isoteric replacement. Tailoring also includes various configurational and stereochemical changes on drug molecule which afford flexibility and overall dimension of drug molecule, for example ring fusion, ring fission, formation of higher or lower homologue, introduction of optically active centre, formation of double bond towards geometrical isomerism and introduction of bulky group towards restricted rotation etc. It involves the study of effects of biologically active compounds on the basis of molecular interaction in terms of molecular structure or its physico-chemical properties involved. It studies the processes by which the drug produce their effect, how they react with the protoplasm to elicit the particular pharmacological effects or response, how they are modified or detoxified, metabolized or eliminated by the organism.

Thus, drug design involves either total innovation of lead or an optimization of already available lead. The lead is a prototype compound that has the desired biological or pharmacological activity but may have many undesirable characteristics. So the current trend in drug design is to develop new clinically effective agents through the structural modification of lead nucleus.

The range of chemical compounds is virtually, and it will never be possible to explore it fully. This is particularly true in the search for new therapeutics because the multitude of biological systems and the tests they require have added a new dimension. Therefore, any procedure is unavoidably bound to consist of the selection of subsystems from a large group of compounds. A subsystem is in general understood to represent a number of chemical compounds formed by substantial variations in a given parent structure, referred to as the lead compound. The most important approaches currently employed to obtain sub-system or promising lead compounds are as follows[3]:

1. Mass screening
2. Haphazard discovery
3. Modification of natural compounds
4. Discovery and exploration of side effects
5. Investigation of drug metabolites
6. Chemical modification of natural endogenous substances

All these procedures are rather empirical, and much is being left to accident. The problem becomes even more serious where compounds for unknown action are to be designed.

Another approach of developing a new drug involves the screening of a large number of new compounds of unusual structure for indications of pharmacological response and action against bacteria and

viruses. Currently, screening centres all over the world are seriously engaged in designing and testing newer compounds of miscellaneous structures for their therapeutic value in hypertension, cancer and other diseases. Tests are devised to gather maximal amount of information from the available samples. Thus, it is apparent that traditional approaches to treating various diseases have ranged from natural products to synthetic molecules. These traditional drugs act by interfering with the action of disease associated proteins or enzymes. These include antimetabolites inhibitors, substrates, analogs etc. However, the diseases caused by viruses appear at genetic level and necessarily require inhibition of the appearance of disease associated proteins.

At this stage, appropriate theoretical methods might prove to be of utmost advantage in the development of drugs. Long ago, Crum-Brown and Fraiser[4] laid the foundation of such a theoretical approach proposing that the biological activity of a compound is a function of its chemical properties. With this concept, structure activity relationships (SAR)[5-10] are developed when a set of physico-chemical properties of a group of congeners is found to explain variations in biological response of those compounds. This has resulted in the discovery, examination and interpretation of structure activity relationships in a more systematic way, which has led to the introduction of quantitative structure activity relationships (QSAR).

1.1 MOLECULAR MODELLING

The field of molecular modelling[11] is commonly thought of as being composed of several interlinked activities, including molecular graphics, computational chemistry, statistical modelling and to some degree, molecular data and information management. The molecular graphics aspect represents the drug molecules and their associated molecular properties in a visual way, so that one may gain greater insight

into their pharmacologic behaviour. The computational chemistry compounds is concerned with simulation of atomic and molecular properties of compounds of medicinal interest through equations, and with the numeric methods used to solve these equations on the computer statistical modelling encompasses the search for quantitative relationships between the structures or properties of a series of compounds and their resultant biological activities. This aspect of medicinal chemistry enterprise is called Quantitative Structure Activity Relationship (QSAR). The chemical data/information management, part of the properties of thousands of compounds into an extensive database, capable of being searched for highly promising compounds with the right combination of properties to make them candidates of pharmacologic evaluation. Another component aids the chemist in the synthesis of new drugs by providing strategies and choices of ways to accomplish the organic synthesis of a series of drug candidates yet a third component may help to organize the attendant molecular properties of series of compounds to make them easier to subject to statistical analysis, such as QSAR.

The common component of all these activities is the computer. Thus when all of the aforementioned activities are taken together, they constitute the field of medicinal chemistry known as Compute-Assisted Molecular Design (CAMD) or Computer-Assisted Drug Design (CADD).

1.2 COMPUTER-ASSISTED DRUG DESIGN (CADD)

CADD means:

(a) no recipe for patents

(b) helping in the decision which additional derivatives should be synthesized (optimization)

(c) helping in detection of new leads and of exception.

(d) helping in our understanding of complex processes involved in drug action (mechanism).

(e) helping in testing working hypothesis.

(f) helping to analyze multivariate data from various test systems.

(g) forcing to perform disciplined and quantitative data analysis.

(h) forcing to do far sighted experimental design.

(i) demanding and supporting interdisciplinary co-operation.

(j) etc.

The drug discovery and lead optimization process is currently dominated by developments in two fields[12].

(1) a **'rational design'** based on structural information and sophisticated computer methods to elucidate the structural prerequisite for binding to particular target.

(2) a **'random screening'** using high through-put-screening techniques to discover possible leads from large compound. Libraries provided increasingly by combinatorial chemistry.

The above two approaches are complementary structural characteristics about a particular series of compounds which can be used to establish a SAR. The derived model helps to explain the important relative differences within a compounds series, suggests how to improve their binding properties and assists in ranking and selecting novel candidates for synthesis.

1.2.1 DRUG DISCOVERY AND DEVELOPMENT PROCESS

The process of drug discovery[13] is a long, tedious and expensive one. The steps involved are:

1. New lead discovery
2. Lead optimization
3. Priclinical lead development
4. Clinical lead development
5. Post marketing surveillance

[1] New Lead Discovery

Their are number of ways of discovering new leads :

(i) Isolation of active substances from natural products for example Penicillin.

(ii) Derivation and application of structure activity data for example Cephalosporin.

(iii) Structure directed molecular design for example Carbonic anhydrase inhibitors.

(iv) Chemists's intuition for example Enalapril.

(v) Modification of natural products for example Theinamycin.

(vi) Broad screening of known synthetic compounds for example Sulfa drugs.

[2] Lead Optimization

It can be done by following ways:

(i) Synthesis and testing of congeneric structures.

(ii) Develop structure activity and/or mechanism of action based models.

(iii) Calculate physical properties and correlate them with activity.

[3] Preclinical Lead Development

It consists of following studies:

(i) Drug formulation experiments.

(ii) Dose ranging studies in animals.

(iii) Animal safety studies.

(iv) Drug delivery/elimination/metabolism studies.

(v) Develop large-scale synthesis.

[4] Clinical Development

It requires following steps:

(i) Small-scale safety and dose ranging tests in healthy humans.

(ii) Develop clinical study protocols obtained approval.

(iii) Recruit clinical investigators and patients for study.

(iv) Carry out the study.

(v) Analyze and report results.

[5] Post Marketing Surveillance

It consists of collection of usage and side effects report.

1.2.2 CONTRIBUTIONS AND ACHIEVEMENTS OF CADD

Drug discovered upto now by classical methods were designed by trial and error method. Availability of CAMM systems has created new horizon for the design of new drug molecules.

CADD can contribute not only to the design of potent compounds but also contributes to many steps in the development of a new drug from laboratory to clinic. It helps in discovering new lead structure as well as in lead optimization.

CADD provides-

(i) 3D-structure of molecules.

(ii) Chemical and physical characteristics of the molecule.

(iii) Comparison of the structure of one molecule with other different molecules.

(iv) Visualization of complexes formed between different molecules.

(v) Prediction about how related molecules might look a number of examples now exists that clearly show that CADD had made major contributions to the drug discovery processes[14].

(vi) Design of thymidylate synthetase inhibitors as anticancer agents (1191).

(vii) Design of HIV protease inhibitors as antiviral agents (1994).

(viii) Design of neutrophil elastase inhibitors as an antiglaucome agents (1989).

(ix) Discovery of novel sweeteners using a sweet taste receptor model (1990).

1.2.3 REQUIREMENTS FOR CADD

Their are two main requirements[15] for molecular modelling systems:

(1) Graphics Hardware

The predominant system for molecular modelling calculations are workstations with UNIX operating system. For example 3D-graphics workstation from Silicon Graphics and Evans and Sutherland Multi-picture systems. But the entire range of computer hardware is being used for CADD such as:

(i) Desktop Mackintosh

(ii) MS DOS personal computers

(iii) Computer Servers

(iv) Super computers such as oray super computers.

(2) Software Packages

A variety of commercial packing are available, ranging from $ 50 to $ 500 for PC based systems. Unfortunately, at preset there is no one system that meets all the needs of the molecular modeller. Major currently available and commercial molecular modelling software systems are; Catalyst, Concord, Chem-X, Amber, Frodo, Sybyl/Alchemy, Cerius 2, Apex-3D, etc.

1.2.4 APPROACHES USED IN CADD

Two approaches are mainly used in CADD

(A) Direct drug design

(B) Indirect drug design

(A) Direct Drug Design

In direct drug design[15a], the 3D-features of the receptor site are directly considered for the design of new drug molecules (X-ray structures or 3D-model of an enzyme). The 3D features of a receptor consist of electronic distribution, stereo features and hydrophobic potential. Based on lock and key fit of a drug molecule and the receptor i.e. considering complementarily of structure of the drug and receptor, drug molecule is designed from the known structure of the receptor.

The approach is limited in use due to difficulty in determining the exact structure of the receptors. Since the receptors are biomolecules, and mostly complex in nature, indirect drug design approach is more common in use.

(B) Indirect Drug Design

In indirect drug design[15b], the analysis is based on the comparison of the stereochemical and physicochemical features of a set of known active/inactive molecules and is interpreted in terms of complimentarity with the structure of the unknown receptor site. Structural analysis consideration and differences between the series of biological and synthetic molecules lead to the design of new substances along their path. The commonly used techniques for indirect drug designer.

1. QSAR analysis
2. Molecular shape analysis
3. Receptor surface model generation
4. Phamacophore mapping
5. Comparative Molecular Field Analysis (CoMFA)

General steps involved in applying these techniques are:

(i) Sketching of the molecules

(ii) Energy minimization

(iii) Conformational analysis and geometry optimization of the compounds.

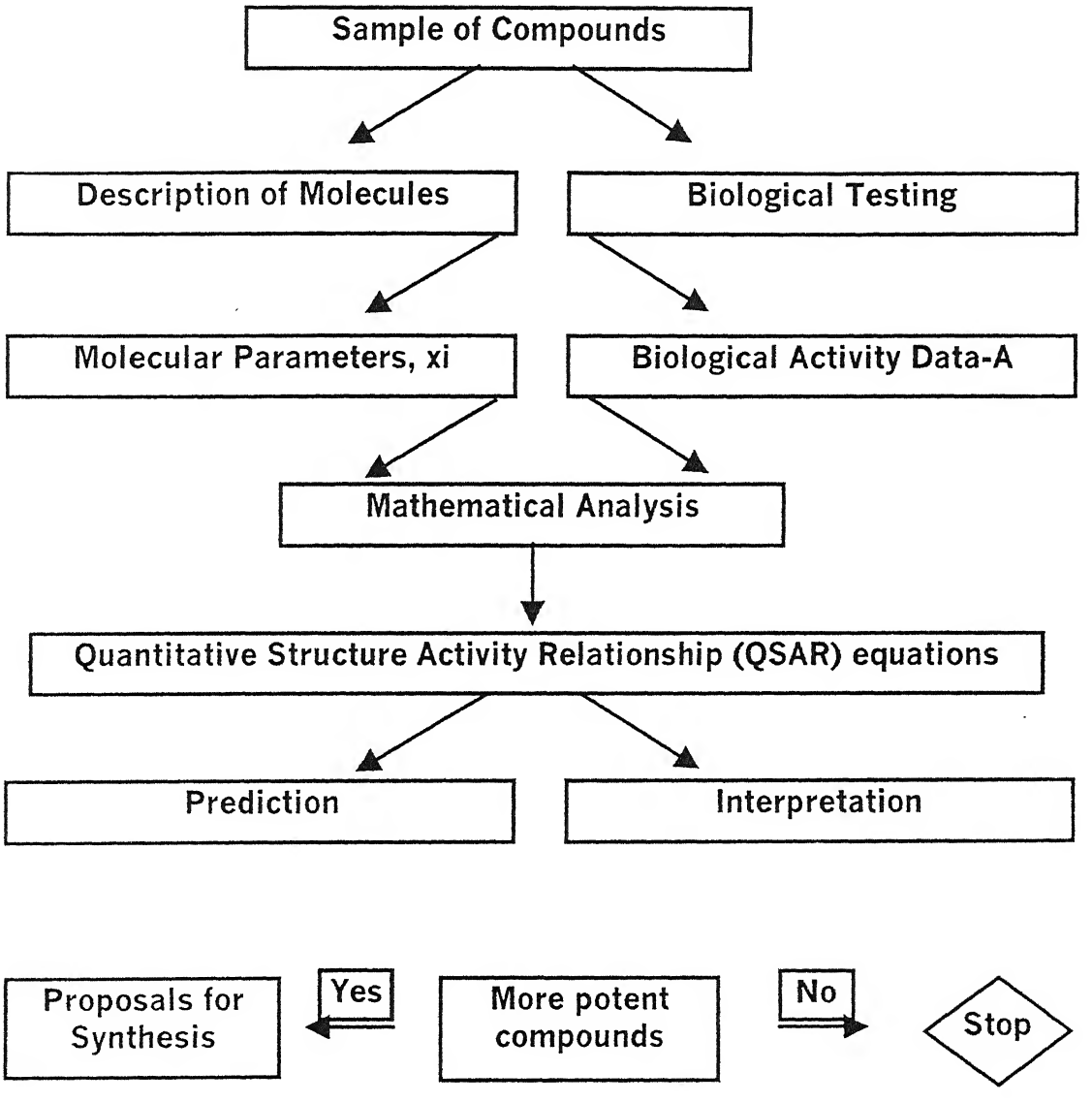
1.3 QSAR ANALYSIS

In medicinal chemistry, the early 1960's marked the surfacing of well defined method of quantitative studies of structure activity relationships QSAR[8-10,16-17] with the publication of linear multiple regression model by Hansch and Fujita, the additive model by Free and Wilson, and the similar interaction model by Bocek and Kopecky. The QSAR methods have not been able to replace the intuitive approach although they have been of aid in reducing the number of educated guesses in molecular modification. Nevertheless, they have contributed directly to the practice of drug design and medicinal research.

The fit of the structure and the complimentarity of the surface properties of a drug to its binding sites at the receptor/enzyme is essential for its biological activity. QSAR helps to understand structure activity relationships in a quantitative manner and to find the influence of certain properties on the biological activity and the strategy enables chemists to look at their structure also in terms of their physicochemical properties in addition. Now a days 3D molecular modelling enables a chemist two maintain a structure bases storing 3D structures, quantumchemical and physicochemical parameters of chemical compounds and activity data. It also helps him with an automated generation of knowledge bases storing quantitative and qualitative structure activity relationships represented by rules and models and also in prediction of biological activity of novel compounds.

Most often QSAR analyses are retrospective studies, whether they follow a rational design of investigated structures or not. Only after performing synthesis and biological testing, a quantitative relationship is derived. Often the optimization of a lead compound is step by step accompanied by QSAR analysis.

The "**Principle of QSAR**" may be illustrated as[18-22]



1.3.1 PRINCIPLE OF QUANTITATIVE STRUCTURE ACTIVITY ANALYSIS

The resulting biological activity parameters, "A" and molecular parameters, "Xi" are related, since biological activity is dependent on molecular structure and the resulting properties. Mathematical analysis reveals such connections in the form of so called quantitative structure-activity relationships (QSARs).

QSARs can be constructed for different purposes and according to different methods structure response relationships describe the connections between the magnitude of a given biological effect and the drug structure in a set of congeners, they can therefore be employed to optimize the effect based on structural variations.

The ultimate objective of QSARs is the prediction of either hypothesis on the mechanism of action or new analogues with any of the above methods, QSAR can help in recalling similar structures of biological activity profiles by computer analysis. It can also calculate the significance of the analogs at hand and there by suggest additional analogs to be synthesized. Finally, if a series of analogs shows no discernible trend of increased potency or specificity, QSAR may indicate that the best compound obtainable in this series has already been prepared and that work on these compounds should be terminated. Description of the molecular structure, electronic orbital distribution, reactivity, reaction rates and the role of structural and steric components and constituents of chemical compounds has been the subject of mathematical formulation by physical-organic chemists. Its conclusions were based on physical measurements. Among these are the determination of pKa and lipophilicity. The contribution of QSAR after 1964 was to quantitatize and evaluate relationships between such physical properties and biological activities and to improve the methodology of measurements. For this purpose, the

chemical structure of a compound has to be transformed into a set of numerical descriptors.

QSAR have been used to forecast biological activities with varying degrees of reliability.

A number of success stories are listed in an excellent review by Martin[23]. This review also catalogs the lack of universal success of predicting potency and mentions the following preventable circumstances that led to failure:

1. The prediction was based on a poorly designed series on an ambiguous regression equation.

2. It was based on extrapolations outside the range of the physical properties represented by the original substituents.

3. The conditions of the biological tests were different.

One can expect that such experimental errors will occur less frequently as more medicinal chemist become expert in QSAR and that the reliability of the method will improve.

It must always be remembered that prediction derived from QSARs have a statistical character and thus always a certain probability of being in error. A basic limitation of structure activity analysis lies in the fact that only such information can be extracted as is present in the biological data. Thus structure activity analysis by itself can not lead to new concepts in therapy, although the manipulation of the information contained in biological data and chemical structures through the use of large computers and appropriate programs may also aid in developing new concepts.

In summary, it may be said that methods of quantitative structure activity analysis in their current stage of development have already been

found to be extremely helpful instruments without which no effective search for new drug is possible and which will undoubtedly offer further possibilities in both drug research and elucidation of biological mechanisms.[24-35]

1.4 DRUG RECEPTOR INTERACTIONS

The concept of interaction of drugs with certain substances with which they are capable of forming compounds according to their chemical affinity goes back to the work of Langley in 1873-78[36]. The stereospecificity of such interactions were recognized by Fischer in 1894[37]. In the following study **"receptor"** is used as a synonym for any biological target for example any specific binding site of a macromolecule, strictly speaking this broad meaning is not correct from ours today's definition of receptors as being soluble membrane anchored or membrane embedded proteins that are able to produce certain biological response via a series of mostly unknown events[38-40].

During the past decades the originally static lock and key mode of ligand-receptor interaction was modified to a more realistic picture, with flexible drug molecules and dynamic receptors[41,42]. Whenever a ligand approaches its binding site, both partners may change their shape (induced fit, flexible fit). Most of our knowledge regarding the geometry of ligand binding site interactions resulted from 3D structures of soluble proteins, especially of enzymes and their inhibitor complexes[43-45].

An important contribution to the receptor concept resulted from recent investigations of Herbette [46,47] of the partitioning into and the distribution of drugs in biological membranes. The correct spatial arrangement of the drug and its proper orientation in the membrane w.r.t. the binding site at the surface of the membrane embedded receptor are considered to be of atmost importance for the drug receptor interaction (Fig. I, II, III).

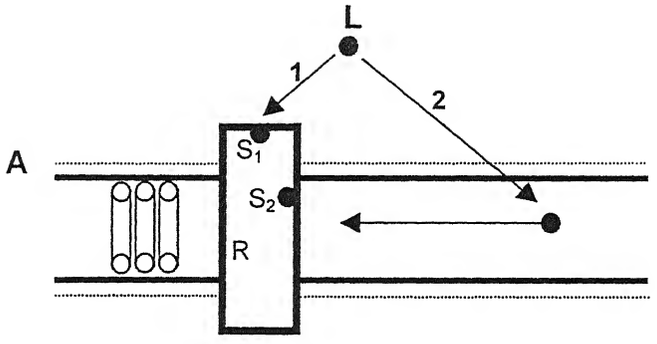


Figure - I : A ligand may reach its binding site S_1 or S_2 at the receptor R by direct diffusion in the aqueous medium as in the case of site S_2 by partitioning into the membrane and then diffusing to the binding site.

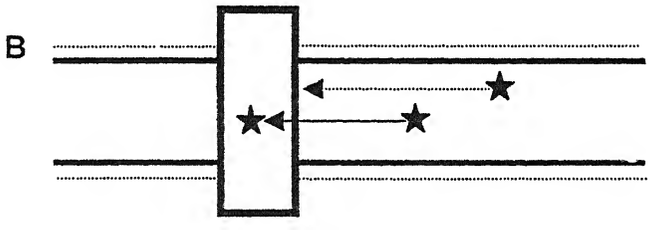


Figure - II : The highly ordered structure of the lipid bilayer may restraint lipophilic drugs to a particular depth of penetration so it becomes important for the ligand to reach on optimum depth in order to show a maximum activity.

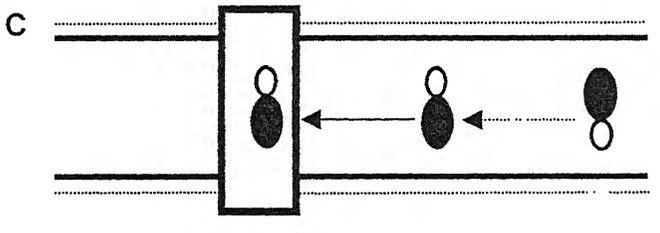


Figure - III : The orientation of the ligand relative to the binding site might also be optimized by the membranes by limiting the rotational degree of freedom of the drugs. So it becomes necessary for the chemist to design the molecule in order to make it orientation specific.

Here,

A ⟶ Position specific
B ⟶ Target specific
C ⟶ Orientation specific

SECTION-B

QSAR METHODOLOGY AND PARAMETRIZATION

QSAR is nothing but the prediction of biological activity of a molecule prior to evaluation or even synthesis in order to reduce the costly and time consuming synthetic work and biological screening[48].

Now medicinal chemist can achieve more potent drugs, occurring to the development of large matrix of QSAR methodology, with lesser dependence on trial and error synthesis and testing.

2.1 HISTORY AND DEVELOPMENT OF QSAR

In 1868, Crum-Brown and Fraiser[49-56] published an equation, which is considered to be the first general formulation of a QSAR. They observed that distinct changes in the degree of activity paralleling somewhat minor changes in chemical structures. Therefore they assumed that the **"physiological activity"** ϕ must be a function of the chemical structure **C**

$$\phi = f(C) \qquad \rightarrow (1)$$

In 1893, Richet concluded that the degree of activity of organic compounds is inversely related to their water solubility. This postulate is known as Richet's rule.

$$\Delta\phi = f(\Delta C) \qquad \rightarrow (2)$$

where, $\Delta\phi$ = differences in biological activity values.

ΔC = corresponding changes in the chemical and physico-chemical properties.

At the turn of twentieth century, Meyer and Overton independently observed linear relationship between lipophilicity, expressed as oil water partition coefficients and narcotic activities.

According to Fuhner, with homologous series, narcotic activities increase in a geometric progression i.e. $1:3:3^2:3^3$ etc., which gave the first evidence of an additivity of group contributions to biological activity values.

In 1939, Ferguson observed a **"Cut Off"** biological activities beyond a certain range of lipophilicity and gave thermodynamic interpretation for such non linear structure activity relationship.

In 1956, Bruice, Kharasch and Winzler formulated group contributions to biological activity values in a series of thyroid hormone analogs. Using equation (3), they obtained excellent correlations between calculated and observed biological activities:

$$\textit{log \% thyroxine - like activity} = K\Sigma f + C \qquad \rightarrow (3)$$

where $\Sigma f = (f_x + f_x^{\,l} + f_{OR}^{\,l})$. Subscripts x, x^l or OR^l represent substituent positions of the molecule and C is the constant. In 1964, an important contribution in the development of QSAR model was made by Free and Wilson (**"de novo"**) model. They defined the biological response (BR) as equal to the sum of the contributions to the activity of the substituent groups plus the overall average activity (μ) which might be attributed to the activity contribution of the parent structure as given in equation (3).

$$BR = (\textit{substituent group contribution}) + \mu \qquad \rightarrow (4)$$

At the same time, Hansch and Fujita developed approach called Hansch analysis (or linear free energy related approach, LFER, or extrathermodynamic approach).

2.2 DEMANDS ON BIOLOGICAL ACTIVITY DATA USED IN QSAR ANALYSIS

The various demands that should be mate for biological activity data used in QSAR Analysis[57] are as follows:

(1) Large range in observed activities

(2) Identical mode of action (parallel dose response curve)

(3) Concentration in molar units (does in g/kg is not suitable)

(4) Activity data as a function of concentration (ED_{50}, LD_{50}, etc.)

(5) Activity data in percentage (protein binding, metabolism, etc. have to be transformed).

(6) Attention, possible time dependency (steady state)

Usually QSAR techniques are divided into two classes[58-59].

- Classical QSAR
- 3D-QSAR

2.3 QUANTITATIVE MODELS

2.3.1 HANSCH MODEL

(The Extrathermodynamic Approach)

Hansch analysis [55, 60-63] correlates biological activity values with physicochemical properties by linear, linear multiple or nonlinear regression analysis, thus, Hansch analysis is indeed a property-property relationship model. As practically all parameters used in Hansch analysis are linear free energy-related values (i.e. derived from rate of equilibrium constants) the terms **"linear free energy related approach"** or **"extrathermodynamic approach"**[64] are sometimes used as synonyms for

Hansch analysis. Also the biological activity values are if they are properly defined, linear free energy-related values (for example binding of inhibition constants, absorption and distribution rate constants or complex data which correspond to a weighted combination of several unit processes).

Hansch proposed the action of drug as depending on two processes. Firstly the journey from the point of entry to the body to the site of action and secondly the interaction with the receptor site. He suggested the linear and non-linear dependence of biological activity on different parameters.

Linear -

$$log\left(\frac{I}{c}\right) = a\,log\,P + b\sigma + cEs + d \qquad \rightarrow (5)$$

Non-linear

$$log\left(\frac{I}{c}\right) = a\,log\,P + b(log\,P)^2 \pm (c\sigma \pm dEs \pm e) \qquad \rightarrow (6)$$

Key observations are :-

(1) Biological activity should be quantitatively related to a set of theoretical parameters, which are assumed to describe essential properties of the drug molecule.

(2) The coefficients are determined by multiple linear regression analysis.

(3) Activity of a drug is controlled by various additive factors.

(4) Biological activity could be described by more than one parameter. So multiple linear regression could be used for drug design.

2.3.1.1 ADVANTAGES AND DISADVANTAGES OF HANSCH ANALYSIS

Advantages:

(1) Use of descriptors (pi, sigma, E_s, etc.) from small organic molecules may be applied to biological systems.

(2) Predictions are quantitative and may be evaluated statistically.

(3) It is quick and easy.

(4) Potential extrapolation conclusions reached may be extended to chemical substituents not included in the original analysis.

Disadvantages:

(1) Descriptors are required for substituents being used.

(2) Large number of compounds required (training set for which physicochemical parameters and biological activity is available).

(3) Limitations associated with using small molecule descriptors on biological systems.

(4) Steric factors of limited applicability in biological systems.

(5) Predictions limited to structural class. (congeneric series)

(6) Extrapolations beyond the values of descriptors used in the study are limited.

2.3.2 FREE WILSON MODEL (THE ADDITIVITY MODEL)

(de novo approach)

This method is based on the assumption that the introduction of a particular substituent at a particular molecular position always leads to quantitatively similar effect on biological potency of the whole molecule. The Free Wilson approach is a true structure-activity relationship model. An indicator variable is generated for each structural feature that deviates from an arbitrary chosen compound values one, indicating the presence of a certain substituent or structural feature, and zero, indicating its absence are correlated with the biological activity values by linear multiple regression analysis. The resulting regression coefficients of the indicator variable are the biological activity contributions of the corresponding structural elements; **"Mathematical model" "Additivity model"** or **"de novo approach"** are synonyms for the Free Wilson method.

It is expressed by the equation -

Log BA = Contribution of unsubstituted parent compound + contribution of corresponding substituent.

$$Log\ BA = \mu + \Sigma(aij) \qquad \rightarrow (7)$$

where, 'i' is the number of the position of which substitution occurs and 'j' is the number of the substituents at that position.

The basic assumptions for the use of Free Wilson method are :

(1) All drug tested should have the same parent structure.

(2) The substitution pattern in various derivatives has to be same.

(3) The substituents have to contribute to the biological activity additively and in the same position with a constant amount being independent of the presence or absence of the other substituents in the molecule.

A major limitation of this approach is that it is not reversible (like π, σ etc. It is applicable only to the system under study.

2.3.3 THE MIXED APRROACH

(The Relationship between Hansch and Free Wilson Analysis)

Hansch analysis and the Free Wilson method differ in their application, but they are never the less closely related[65, 66,67]. Kubinyi has presented the contribution of Hansch and Free Wilson models as **"mixed approach"**. The mixed approach can be written as :

$$log\frac{1}{C} = \Sigma\ aij + \Sigma\ Kj\phi j + K \qquad \rightarrow (8)$$

Where, 'Kj' represents the coefficient of different physiochemical parameters.

In this equation, 'Σaij' is the Free Wilson part for the substituents and $\phi j = \pi$, σ and Es contribution of the parent skelton. The mixed approach was developed to find possible interactions between Free Wilson parameters and physico-chemical properties of the substitutes used.

The basic assumptions of this method are -

(1) All the drug tested have some parent structure.

(2) The substituent pattern in various derivatives has to be the same.

(3) The substituents contribution to the biological activity additively being independent of the presence or absence of other substituents.

Today mixed approach is the most powerful tool for the quantitative description of large and structurally diverse data sets.

2.3.4 OTHER QSAR APPROACHES

2.3.4.1 CRAMER'S SUB -STRUCTURAL ANALYSIS

Berkoff, Crammer[68-70] and Redl developed this method which is one of the first attempts to apply substructural analysis in the QSAR field. The compounds of a training series are fragmented into structures using a library of atom, bond and sub structural topological features. For each features sub structural activity frequency (SAF) is calculated as:

$$SAF_j = \frac{\text{number of active. compounds containing the feature j}}{\text{total number of compounds containing the feature j.}} \rightarrow (9)$$

Where SAF_j is the probability contribution of the j^{th} feature to the overall probability to characterize the activity of compounds the mean substructure activity frequency for i^{th} compound is given as

$$MSAF_i = \frac{1}{m_i} \Sigma b_{ij} SAF_j \rightarrow (10)$$

where m_i is the number of features (fragments) occurring in the i^{th} compound and b_{ij} is a substructural descriptor defined as:

$$b_{ij} = \begin{cases} 1, \text{if the } J^{th} \text{ feature is present in the } i^{th} \text{ compound.} \\ 0, \text{if not} \end{cases} \rightarrow (11)$$

On the basis of experimental studies, it is observed that MSAF values are related with biological activity.

2.3.4.2 PRINCIPAL COMPONENT ANALYSIS

In Principal Component Analysis, data matrices obtained by measuring a given set of variables for a given set of objects are examined. Variables are presented as linear combinations of new variables called principal components.

This analysis is basically a mathematical method of describing and reproducing variables of data matrix by means of a new set of "**abstract**" variables, i.e., the principal components.

This approach has extensively been discussed by Weiner and Malinowski who applied it to a variety of problems in chemistry.[71-78]

The model of principal component analysis may be written as:

$$x_{ij} = \sum_{k=1}^{r} u_{ik}\, v_{kj} + x_j + E_{ij} \qquad \rightarrow (12)$$

or in matrix notation-

$$X = UV + \bar{X} + E \qquad \rightarrow (13)$$

Where:

x_{ij} = value of the j^{th} variable for the i^{th} object.

r = minimum number of components necessary to reproduce the x_{ij} within E_{ij}.

u_{jk} = i^{th} element of the k^{th} so called object component describing the k^{th} property of the i^{th} object.

v_{kj} = j^{th} element of the k^{th} so called system component describing the k^{th} property of the j^{th} variable.

x_j = mean value of the j^{th} variable

E_{ij} = residual error)experimental plus model error).

X = data matrix with the objects in the rows and the variables in the columns.

U = object component matrix (elements : u_{jk})

V = system component matrix (elements : v_{kj})

X = vector of mean values (elements : x_j)

E = error matrix (elements : E_{ij}).

Once all object components have been identified with test vectors they can be replaced by the later. Equation (12) then transforms to:

$$X_{ij} = \Sigma u_{jk} + V_{kj} + C_j + E_{ij} \qquad \rightarrow (14)$$

equation (14) is the desired result of principal component analysis, representing a multivariate system of **"regression"** equations.

This method requires data of sufficient precision and objectively existing relations among the various systems.

The main steps of principal component analysis are summarized as:

1. Calculation of the correlation matrix.
2. Evaluation of object and system components by the principal component method.
3. Determination of the (r) of relevant principal components necessary to reproduce the data matrix within experimental error.
4. Uniqueness test.
5. Selection of test vectors.
6. Identification of the object components with test vectors.
7. Formulation of the system of "regression" equations describing the variables in terms of the parameters used as test vectors.

2.3.4.3RANK CORRELATION ANALYSIS.

Sklenar and Jager[79-80] developed this analysis.

Before applying this procedure to QSAR[39-41], it is assumed that within homologous series there exists a defined and monotonous relation between the level of biological action and molecular properties relevant to

that action. This is less strict requirement than the linear model assume to hold in the extra thermodynamic approach which is certainly realistic especially when dealing with quantum chemical molecular parameters. Such relations can be examined by means of rank correlation analysis. The first step comprises the transformation of the values of biological activity and the molecular parameters concerned into rank numbers. These numbers indicate the position of each value when the respective data are ranked in decreasing order.

Width the aid of the rank numbers and equation-

$$r_{sj} = 1 - 6\sum_{i=1}^{n} d_i^2 / (n^3 - n) \qquad \rightarrow (15)$$

where, d_i denotes the difference of the rank numbers of parameter x_j and biological activity of the i^{th} compound and n is the number of compounds in the sample. This equation (15) serves to characterise the connection between molecular property (given) and biological activity. If the correlation between A and x_j is perfect = $|r_{sj}| = 1$.

2.3.4.4LINEAR DISCRIMINANT ANALYSIS

The discriminant functions in this analysis serve as classification algorithms, representing weighted linear combinations of features relevant for class separation. There are several types of discriminant functions, among which the non-elementary ones have optimal properties.

If Q classes occur (q = 1 .....Q), Q - 1 non elementary discriminant function w_k (k = 1.....Q - 1) of the general form

$$w_k = \sum_{j=1}^{m} a_{kj} x_j \qquad \rightarrow (16)$$

where, a_{kj} denotes the weight factor (coefficient) of the j^{th} molecular parameter (j = 1 .....m) in the k^{th} discriminant function and the w_k's being designated as non-elementary discriminant variables.

The main steps of this analysis are:

1. Selection of molecular parameters and formulation of approaches.
2. Check whether classes can be separated using these parameters and elimination of parameters not relevant for class separation.
3. Computation of non elementary discriminant functions.
4. Interpretation of discriminant functions and reclassification of the compounds from the training series as a means of testing the quality of separation.
5. Classification of compounds not yet investigated so that further synthesis and tests can be planned while allowing for mechanistic conclusions from the form of the discriminant functions.

Using several modifications discriminant analysis has been widely and successfully applied in the QSAR field by several workers[81-88].

2.3.4.5 PATTERN RECOGNITION TECHNIQUES

Pattern recognition technique[89-90] is similar to the classical QSAR method, only the number of variables is much larger in this than in Hansch analysis. Problems are associated with proper selection of a training set and stepwise regression. Therefore multivariate methods like PCA or soft modelling techniques for example SIMCA or PLS analysis are used instead of regression analysis.

2.3.4.6 ARTIFICIAL INTELLIGENCE PROGRAMS

CASE (Computer Automated Structure Evaluation)[91,92] automatically identifies molecular features (biophores) that contribute to biological activity. A machine learning program GOLEM from the field of

inductive logic programming uses activity, structural and stereochemistry information of active and inactive analysis to derive inductive hypothesis.

2.3.4.7 MOLECULAR ORBITAL METHOD

Molecular orbital method[93] method is one of the most important among several semi-empirical approaches used in drug design. As per the basic assumption of this theory the electrons are considered as being associated with molecule as a whole rather than with a particular substructure. The molecular orbital wave functions (φ) constructed from atomic orbital provides information about the physical properties of a molecule. This is turn helps to get information about ionization potential, electron affinity etc.

2.4 3-D QSAR

Three-Dimensional Quantitative Structure Activity Relationship (3D-QSAR)[94] analysis is a sufficiently new area of Computer Assisted Molecular Design (CAMD). 3D-QSAR study is an ability to characterize the shape and charge distributions of a molecule in 3D shape. This is a consequence of the predominance of steric and electrostatic interactions in the binding of a drug to its targeted receptor site.

3D-QSAR are developed for specific, highly anisotropic ligand-receptor interactions that correspond to **in vitro** biological assays. In the very limited applications of 3D-QSAR to **in vivo** sets, the problem has usually been thought of in terms of a specific ligand receptor interaction component (specific 3D descriptors) and a non-specific transport metabolism or other, component (general-thermodynamic descriptors such as log P). Overall, 3D-QSARs probe and extract information about a specific interaction involving the ligand, which almost always, also involves the ligand's receptor site. Thus, one useful way to compare and contrast 3D-QSAR methods is to identify what aspects of the general ligand-receptor-binding process are being considered in a particular 3D-QSAR formalism.

2.5 PHARMACOPHORE MAPPING

A phamacophore or phamacophore pattern is the set of features required for a compound to elicit certain biological activity. These features are typically any combination of structural, chemical and physical attributes of a molecular structure. The process of developing a pharmacophoric pattern i.e. pharmacophore mapping[95] involves three main aspects:

1. Finding the features required for biological activity.

2. Determining the molecular conformation required (i.e. the bioactive conformation)

3. Developing a superposition as alignment rule for a series of compounds.

2.6 3D-QSAR APPROACHES

It deals with the various techniques of pharmacophore mapping which serve the important approaches and are as following:

2.6.1 COMPARATIVE MOLECULAR FIELD ANALYSIS (CoMFA)

To some 3D-QSAR analysis and CoMFA[94,96] are one and the same. CoMFA is by far the most often employed R1 3D-QSAR approach, reflecting a novel, conceptually satisfying scientific approach reduced to practice as a well-written and versatile software package. There are many reports in the literature of successful application of CoMFA that have not only led to predictive models within an analogue series of biologically active molecules[97] but also to insightful information on the general requirements for the expression of the activity.

It is a 3D-QSAR approach which places molecules in a grid and correlates a field properties (electronic and steric field energy) with biological activities.

The method was developed by Cramer et.al. in 1988. It involves the following steps.

1. Selection of reference compound and structural alignment.

2. Electronic charge calculation on each compound of the series.

3. Calculation of electrostatic and static field energies at various grid points in a lattice of specified dimensions using probe atoms or groups.

4. Regression analysis of specified dimensions using probe atoms or groups.

5. Testing of CoMFA models.

Because of the increasing availability of powerful computational hardware and software and of the advances in our basic understanding of theoretical chemistry, the chemists now have possibility of calculating many properties of molecule which have not yet been synthesized and of being reasonably assumed that the synthetic molecule infact exhibit those properties.

2.6.2 MOLECUALR SHAPE ANALYSIS (MSA)

A formalism that deals with the quantitative characterization, representation and manipulation of molecular shape in the construction of a QSAR is molecular shape analysis (MSA)[98]. This method was developed by Hopfinger. He incorporated conformational analysis to Hansch analysis[99]. The **Common Overlap Steric Volume (COSV)** between a pair of superimposed molecules can be used as a global measure of molecular shape similarity in constructing QSARs[98]. Subsequently, the spatially integrated potential energy field was shown to be a complementary extension of COSV as a general QSAR shape descriptor.

2.6.3 RECEPTOR SURFACE MODEL GENERATION (RSM)

Receptor surface is generated complementary[100] to the active molecules. From this, it can be seen that at which sites interaction with drug takes place in terms of hydrophobic, electrostatic, hydrogen bond acceptor or donor or charge interaction. RSM can be used to predict the activity of new designed compounds. Steps followed are:

1. Selection of reference compounds and structure alignment.
2. RSM generation using van-der-Waal's field function.
3. Mapping of electrostatic potential charge, hydrogen bonding and hydrophobicity properties of the RSM.
4. Evaluation of RSM and calculation of different types of interaction energies between structure and RSM.
5. Generation of QSAR equation.

2.6.4 HYPOTHETICAL ACTIVE SITE LATTICE MODEL (HASL)

The hypothetical active site lattice approach (HASL) [101] is related to the CoMFA methodology and also to MSA. The HASL approach represents each of the shapes of the molecules as a collection of 3D grid points, which is termed the molecular lattice. The resolution of the HASL (i.e. the distance between the grid points) determines the number of lattice points that represent a molecule and also the resolution of the generated receptor map. User defined conformations are selected to generate HASL. Typically, conformations similar in shape are chosen.

The two aims of HASL approach are the prediction of activities of interested compounds, as well as the identification of substructures that most influence the observed activities.

2.6.5 CATALYST

Catalyst is a popular, turn key commercial software package that establishes 3D-QSARs based on a training set of compounds and their activities against a common end point[102]. The turn key aspect of the package refers to the user having to only sketch in the structures in the training set, input the corresponding activity measures and provide some control data, such as the number of confirmation to be sampled in the conformational search for the active conformation. Here graphic representation of the most active compound in its active confirmation with the 3D-pharmacophore is found in the analysis superimposed on the compound.

2.6.6 APEX-3D PROGRAM

Apex-3D program[103] can evaluate 2D (Topological) or 3D (Topographical) relation between the pharmacophoric points. Generation of biophores (pharmacophores) involve determining low energy, representative conformers for each compound, calculation of descriptors for potential biophoric atoms and searching (using a clique detection algorithm) for maximal common 2D or 3D arrangements of biophoric centres. The arrangement or patterns are potential biophores and are then evaluated for their statistical quantitative correlation with biological activity.

2.7 APPLICATION OF QSAR

QSAR has three essential applications[104] :

(1) As an instrument for prediction.

(a) Estimation of physico-chemical properties using substituent constants.

(b) Reduction of the number of compounds to be synthesized.

(c) Faster detection of the most active compound.

(d) Avoidance of synthesis of compounds with some activity.

(2) As a diagnostic instrument

(a) Information on possible types of interaction forces

(b) Information on the **"nature"** of receptor.

(c) Information on the mechanism of action.

(3) Detection of exceptions (outliers).

2.8 LIMITATIONS OF QSAR

Though QSAR studies can be successfully utilized to predict the activity of new analogues and in the determination of mechanism of drug receptor interactions, they have some drawbacks and limitations[105].

The most serious problem in QSAR is the lack of fundamental understanding of how to quantitatively describe substituent effects on non-covalent intermolecular (for example drug-receptor) interactions.

Mutual conformational adaptation of drug and receptor may also occur after interactions. Since no specific parameter has yet been developed for the description of variation in confirmation and conformational flexibility, it imposes limitation on the success of QSAR analysis.

Electronic effects of the substituent may change both the degree of utilization and the charge distribution. The former may affect the amount of active species available to the receptor while the latter may affect the strength of the drug receptor interaction.

QSAR study may be incorrectly interpreted if the biological property of interest is most correctly measured.

2.9 QSAR DESCRIPTORS (QSAR PARAMETERS)

Descriptors are needed to describe the intermolecular forces of the drug receptor interaction, the transport and distribution of drug in quantitative manner and to correlate them with the biological activities. The QSAR descriptors have been broadly classified into :

(1) Conventional and

(2) Non-conventional

Conventional descriptors are used in classical approach towards QSAR. This includes thermodynamic, electronic and steric parameters. Non-conventional or advanced parameters explain the 3D electronic and steric characteristic of the molecules. Examples include molecular volume, molecular surface area, density, dipole moments etc.

Some of the important parameters are described below :

2.9.1 LIPOPHILIC/ THERMODYNAMIC/ HYDROPHOBIC PARAMETERS

No other physiochemical property has attracted as much interest in QSAR studies as lipophilicity (synonymously called by hydrophobicity)[106, 107], due to its direct relationship to solubility in aqueous phases to membrane permeability, and to its (merely entropic) contribution to ligand binding at the receptor site. Some important lipophilic parameters are log P[108] (partition coefficient), R_M (chromatographic parameter) [104], S (Entropy) and π (Hydrophobic constant).

2.9.2 STERIC/SPATIAL PARAMETERS

Steric substitution constant[48,93,109,110] is a measure of bulkiness of group it represent and its effect on the closeness of contact between the drug and the receptor site. As the bulky substituent delays the detachment

of drug from the receptor, it leads to late onset and long duration of action. Steric effects are difficult to describe due to the fact that the 3D structures of the binding sites of drugs are most often unknown. Some important steric parameters are Es (Taft steric Parameter)[93], χ (Molecular Connectivities), MSD (Minimal Steric Difference), L, B_1-B_4 (Sterimol Parameters), PMI (Principal Moment of Inertia), MR (Molar Refractivity), PSA (Polar Surface Area), Vw (van der Waal's volume), etc.

2.9.3 ELECTRONIC PARAMETERS[48, 93, 109, 110]

Electronic parameters describe the influence of certain group or substituent on electron density distribution and thus its effect on biological activity. They affect the metabolism and elimination pattern of the drug and drug receptor interaction. Some of the widely used electronic parameters are : Hammet σ constants[111], Field and Resonance parameters, F and R, parameters derived from molecular spectroscopy, pKa values[112], charge transfer constants, dipole moments, hydrogen bonding parameters and parameters derived from quantum-chemical calculations[113], for example orbital energies and partial charges.

2.9.4 POLARIZABILITY PARAMETERS

Molar volume (MV), molar refractivity (MR), molar polarization (MP) and parachor (PA) are theoretically and practically closely interrelated parameters[114, 115].

$$MV = MW/d \qquad \rightarrow (17)$$

$$MR = \frac{MV(\eta^2-1)}{(\eta^2+2)} = \frac{MW(\eta^2-1)}{d(\eta^2+2)}\left(cm^3/mol\right) \rightarrow (18)$$

$$PM = MV\frac{\Sigma^2-1}{\Sigma^2+2} \qquad \rightarrow (19)$$

$$PA = MV\gamma^{1/4} \quad \rightarrow (20)$$

where,

MW	=	Molecular weight
d	=	density
η	=	refractive index at 20°C
Σ	=	dielectric constant
γ	=	surface tension

Molar volume itself is not strictly additive parameter but the corrected volume parameters, MR and PA are additive constitutive molecular properties. For liquids, the MR value can be calculated in units of volume using Lorentz-Lorentz equation ( ).

MR has been correlated with lipophilicity, molar volume and steric bulk. Due to its MV component, it is related to volume and size of a substituent and thus contributes steric properties. The refractive index related correction term (η) in MR accounts for the polarizability and thus for the size and the polarity of a certain group. Larger the polar part of a molecule, the larger its value will be.

The parachor is a molar volume, MV, which has been corrected for forces of intermolecular attraction by multiplying with the fourth root of surface tension, γ. The parachor has an advantage as steric parameter in that it is easy to calculate either from atomic contributions or from the component chemical bonds.

2.9.5 TOPOLOGICAL PARAMETERS

Numbers reflecting certain structural features of organic molecules that are obtained from the respective molecular graphs are usually called **"topological indices"**. Such a number is usually obtained by imposing certain conditions on vertices (atoms), edges (bonds) or both.

A plethora of topological indices has been considered in the chemical literature and some of them have been found to possess quite remarkable applications in chemistry and in drug research. A topological index expresses topological information for a given chemical structure.

The advantage of topological indices is that they may be used directly as simple numerical descriptors in quantitative structure property activity relationships (QSPR, QSAR). These relationships are mathematical models that enable the prediction of properties and/or activities from structural parameters.

Most of the topological indices are derived either from distance matrix, or adjacency matrix or some combination of both distance as well as adjacency matrix.

Some of the popular topological indices are:

Wiener index (W), Hararay index (Har), Balaban index (B), information theoretic index, (ISIZ), Quadratic index (Qindex), Ramification index (RAM), Kier and Hall valency connectivity indices $\left({}^{m}X_{R}^{v}\right)$, Centralization (Cent), Variation (Var), etc.

2.10 PARAMETERS USED IN THE PRESENT STUDY (THESIS)

1. Log P[108, 116]

It is a free energy related parameter (LFER), which expresses the relative free energy change occurring on moving a substituent from one phase to another. This is an additive property. It means, with the help of π values of the substituents, the log P value of any molecule may be calculated by simple addition.

$$log\ P = \sum_{i}^{m} \pi \text{ (additive free energy)} \qquad \rightarrow (21)$$

Similarly, the hydrophobic substituent, π, of a given substituent X is the difference of log P values of the substituted compound R-X and the unsubstituted compound R-H.

$$log\ P = log\ P_{(P-X)} - log\ P_{(R-H)} \quad \rightarrow (22)$$

$$\pi = log\left(\frac{P_x}{P_H}\right) \quad \rightarrow (23)$$

This parameter describes the permitting behaviour of the molecule in aqueous and lipid phases.

In the present thesis log P is calculated by the method proposed by Moriguchi et. al[117].

2. Molar Refractivity (MR)[109]

This parameter was proposed by Pauling and Pressman. It is a parameter for correlation of dispersion forces in the binding of haptens to antibodies.

It is formulated as:

$$MR = \frac{(\eta^2 - 1)}{\eta^2 + 1)} \cdot \frac{MW}{d} \quad \rightarrow (24)$$

where,

η = refractive index of the compound

d = density of a compound

MW = Molecular weight of the compound

MR is an additive constitutive molecular property, like log P. MR has been correlated with lipophilicity, molar volume and steric bulk. The refractive index-related correction term in MR accounts for polarizability

and thus for the size and polarity of certain groups[118]. The larger the polar part of a molecule is, the larger its MR value will be. A positive sign of MR in a QSAR equation can be explained by binding of the substituents to a polar surface, while a negative sign or a non linear relationship indicates a limited area or steric hindrance at this binding site. In the present thesis MR is calculated by the method proposed by Ghose and Crippen[119, 120].

3. Equalized Electronegativity (x_{eq})

A significant development in the electronegativity concept has been provided by Sanderson's formulation of the principle of electronegativity equalization, which states that "when two or more elements initially different in electronegativity combine chemically, they become adjusted to the same intermediate electronegativity within the compound". This principle which has gained wide acceptance in recent year abandons the idea of fixed electronegativity and redefines the values in electronegativity table as quantities characteristic of isolated atom before a bond is formed[121, 122].

The physical and chemical properties of substances are largely determined by partial charge on the constituent atom[123], and the evaluation of these partial charge is an important electronegativity application. In the framework of Sanderson's principle, it is generally believed that partial charge acquired by an atom through chemical combination is proportional to the difference between the final equalized electronegativity and the initial, pre-bonded electronegativity.

Charge conservation equation leads to a general expression for X_{eq}.

$$X_{eq} = \frac{N}{\Sigma \frac{\nu}{x}} \qquad \rightarrow (25)$$

where,

$N=\Sigma(v)$ = the total number of atoms present in the species formula.

V = Number of atoms of a particular elements in the species formula.

X = Electronegativity of that particular atom.

The group electronegativity is calculated by :

$$X_G = \frac{N_G}{\Sigma \frac{v}{x}} \qquad \rightarrow (26)$$

where, N_G is the of atoms in the group formula.

4. Van der Waal's Volume (V_W)

The Van der Waal's volume (Vw) has been found to be one of the most fundamental characteristics of the drug structure, controlling the biological activity. This determines the molecular size and shape of the compounds, which is very important aspect of drug receptor interactions. Not only this, the hydrophobic behaviour of drug molecule has been shown, to be significantly correlated with V_W (Moriguchi et. al)[117]. Consequently the V_W was found to be related with various biological activities of drugs.

To find Van der Waals volume (V_W) of molecules spherical shapes were assumed for all atom according to Bondi[124], because of the absence of generally accepted pear shapes.

Since Van der Waals radii are greater than covalent radii, a correction for sphere overlapping due to covalent bonding between atoms was needed for the calculation of V_W for polyatomic molecules.

V_W = Σ (Sphere volume of atoms) + Σ (Correction value between atoms) $\rightarrow (27)$

SECTION-C

STATISTICAL METHODS USED IN QSAR ANALYSIS

The primary objective of QSAR is to predict the biological activity for new compounds in the test set. The multivariate statistical approach[125-127] is the best possible tool to be used, since it is the best way of utilizing all the information required at the same time. The various multivariate methods have been developed in describing the structure of the available data sets and therefore to predict the behaviour of the new samples.

1.3.1 REGRESSION ANALYSIS

Regression is a measure of average relationship between two or more variables in terms of original units of the data. Regression analysis correlates independent X-variables (for example physico-chemical parameters, indicator variables) with dependent Y-variables (for example biological data). The dependent variables contain error terms, while the independent variable are supposed to contain no such error. Regression analysis is used for estimating or predicting the unknown value of one variable from the known value of other variable. There can be different types of regression analysis.

a. Simple linear regression

A single independent variable is used for each calculation and a set of QSAR equation is generated. Each equation contains one variable from the descriptor set.

b. Multiple linear regressions

It calculates QSAR equations by performing standard multivariable regression calculations using multiple variables in a single

equation. The variable should be independent and minimize the possibility of chance correlation. The number of independent variables cannot be more than one fifth of the number of compounds in the training sets.

c. Stepwise multiple linear regression

This is useful when numbers of independent variables are very high. It calculates QSAR equations by adding one variable at a time and testing each addition for significance. Only variables found to be significant are used in the first QSAR equation. With this method, a correlation matrix is compared.

1.3.2 DISCRIMINANT ANALYSIS

Discriminant analysis[128,129] method was first introduced by Yvonne Martin in 1974. It is an extension of regression analysis. It is a statistical technique which allows exploration of the significance of correlation between a crude activity parameter (the group) and either continuous or discontinuous indicator variables taking the value 1 to 0 according to presence or absence of certain molecular features. Thus it separate objects with different properties for example active and inactive compounds by deriving a linear combination of some other features.

a. ***COMPACT***[130] (Computer Optimized Molecular Parametric Analysis of Chemical Toxicity) is a discriminant analysis approach used to predict toxicities.

b. ***ALS***[131,132] (Adaptive Least Square Analysis) is a modification of discriminant analysis, which separates several activity classes by a single discriminant function.

c. ***ORMUCS***[133] (Ordered Multicategorial Classification using Simplex Technique) is an ALS related approach which applies simplex technique for derivation of discriminant function. Recently a fuzzy version was developed and used in QSAR studies.

d. *SIMCA*[134] (Similarity, Chemistry and Analogy) is a class modeling technique which places objects from P-dimensional space into lower dimension boxes. Discrimination of objects of different classes is possible by deriving separate principle component models for each class. FA, FCA and NIPALS are some other methods used.

1.3.4 PARTIAL LEAST SQUARE ANALYSIS (PLS)

Partial least square analysis (PLS)[135] is the most useful multivariate statistical method. Many, hundreds or even thousands of independent variables can be correlated with one or several dependent variables. Perfect correlation is obtained due to the usually large number of independent variables. Due to complexity of PLS algorithm and availability of computer programmes for regression analysis, it is not much used. But for 3D-QSAR methods like CoMFA, PLS analysis is the method of choice.

1.3.5. CLUSTER SIGNIFICANCE ANALYSIS (CSA)

To evaluate a congeneric series of compounds, if a graph is plotted by taking biological data (for example active/inactive) on Y-axis and physico-chemical parameters on X-axis, sometimes the active compounds tend to cluster in a relatively confined region of the graph. Such clustering suggest that there is a connection between the parameters and biological activity. The advantage of the method[136] is that qualitative or rank-ordered biological data can be used.

1.3.6 VALIDATION OF QSAR

The QSAR gives us information of how changes in the structure of the actual compounds influence their biological activity[137]. This in turn, allows us to (a) modify the structure in improve drug potency, decrease toxicity etc and (b) improve or understanding of the actual biological mechanism.

$$log\ I/C = b_0 + b_1\pi + b_2\sigma + b_3MR + b_4\pi^2\varepsilon \qquad \rightarrow (28)$$

First necessary condition for model validity is that R^2 is close to 1.0 ($R^2 > 0.90$, $r > 0.95$) and s is small, say, smaller than 0.3, if Y = log I/C. However, a large R^2 and small s is not sufficient for model validity because regression models will give closer fit if there are larger the number of parameters and terms in the model.

Recent developments in statistics provide us with a new interesting set of measures of validity that are based on simulating the predictive power of a model. These tools-bootstrapping and cross-validation operate by creating a number of slight modifications of the original data set, estimating parameters from each of these modified data sets and then calculating the variability of the predictions by each of the resulting models.

Cross-validation which is the simplest to apply, creates a number of modified a way that each observation is taken away once and once only. Then one model is developed for each reduced data set and the response value (y) of the deleted observations are predicted from the model. The squared difference between predicted and actual values are added to the **Predictive Residual Sum of Squares (PRESS)**.

PRESS is a good estimate of the real prediction error of the model, provided that the observations were independent. If PRESS is smaller than the sum of squares of the response values (SSY), the models predict better than chance and can be considered **"statistically significant"**. The ratio of PRESS/SSY can be used also to calculate approximate intervals of predictions of new observations (compounds).

To be reasonable QSAR model, PRESS/SSY should be smaller than 0.4, and a value of this ratio than 0.1 indicates an excellent model.

$$Press \quad = \Sigma[yi - yi)^2 / (1 - hii)^2] \qquad \rightarrow (29)$$

where yi and yi are the response (activity) values of observation i (i = 1, 2, ----- n), observed and predicted by the model respectively. The diagonal elements of the "hat" matrix, H, [$H = X (X^{`}X)^{-1}X^{`}$] are denoted by hii. X = (n x p) is the data matrix containing one column for each of the p terms of the model.

Cross-validation does not work well when-

(i) The observations are strongly grouped and hence not independent. With QSAR, this often happens when two or more different types of compounds are put in the same model.

(ii) The second situation occurs when cross-validation is applied after variable selection in stepwise multiple regression.

Cross-validation may be applied to a large data sets which is used to select the model having the highest predictive ability. In cross-validation, many PLS runs are performed in which one or several objects are eliminated from the data set either randomly or in a systematic manner and the excluded objects are predicted by the corresponding model. This is called leave-one-out (LOO) technique.

1.3.7 EVALUATION OF QSAR EQUATION

When the number of variables exceeds three, the results cannot be expressed in the form of graphs or model. Therefore, a regression equation remains the only method of expression, which can be used situation. Terms commonly used in regression analysis are[125,126]-

1. Correlation Coefficient (r)

It is a relative measure of quality of fit of model. Its value depends on the overall variance of data. High value of r ($r > 0.90$) indicates that the statistical significance of the regression equation is high while low value of r indicates that the substituent constant is not important for the process under consideration. If the r value does not decrease significantly

when a particular substituent constant is omitted from the equation, it means that process represented by equation is least affected by the factor symbolized by that particular substituent constant. Correlation coefficients of the two subsets are relatively small while the correlation coefficient derived from the combined set is much larger, due to the increase in overall variance.

2. Square of the Correlation Coefficient (R^2)

It is a measure of explained variance represented as a percentage value i.e. the term explains about percent data represented by that particular equation. for example if r = 0.7 then R^2 = 0.49 or 49% data is accounted by regression of that parameters, still leaving 51% data yet unaccounted. Thus, the value of r can be improved by the inclusion of another parameters. Greater the value of r^2, lesser is the data (verariance) that remains unaccounted by the equation.

3. Standard Deviation (s)

This value gives us an idea about the precision of that equation. Greater the value of 's', larger will be the accuracy with which the expected potency of a new compound may be guessed. It is an absolute measure of quantity of fit. Its value consider the number of objects (n) and the number of variables (k). Therefore 's' depends not only on the quality of fit but also on the number of degrees of freedom (DF).

$$DF = n - k - 1 \qquad \rightarrow (30)$$

Larger the number of objects and smaller the number of variables is, the smaller the standard deviation (s) for a certain value of $\Sigma\Delta^2$. Normally, s should be around 0.3.

$$S^2 = \frac{\Delta^2}{n-k-1} = \frac{(1-R^2)Syy}{n-k-1} \qquad \rightarrow (31)$$

4. Standard Error of the Coefficient

The figure in bracket following the coefficient represents the standard error of the coefficient, which means that if the experiment is repeated, the coefficient should lie between these limits. Higher the standard error, less reliable is the coefficient and there is a less possibility that the variable is represents is related to the biological response.

5. Number of compounds utilized (n)

For a good correlation, large number of compound must be used. The value of r must be assessed with reference to n for example r=0.89 for n=10 is a better correlation than r=0.98 for n=3.

6. F value

It is a measure of statistical significance of regression model. Only F values being larger than the 95% significance limit prove the overall significance of a regression equation.

$$F = \frac{R^2(n-k-1)}{k(1-R^2)} \qquad \rightarrow (32)$$

It evaluates the statistical validity of a particular equation. For a particular equation if, $S_{standard}$ (i.e. 13.74) < $C_{alculated}$ then, the relationship represented by that equation is statistically significant.

CHAPTER-2

BIOACTIVE COMPOUNDS FOR THE PRESENT STUDY-AN OVERVIEW ESTROGEN RECEPTORS

"**Estrogens**" are a family of related molecules that stimulate the development and maintenance of female characteristics and sexual reproduction. The natural estrogens produced by women are steroid molecules, which means that they are derived from a particular type of molecular skeleton containing four rings of carbon atoms, giving the shape shown here.

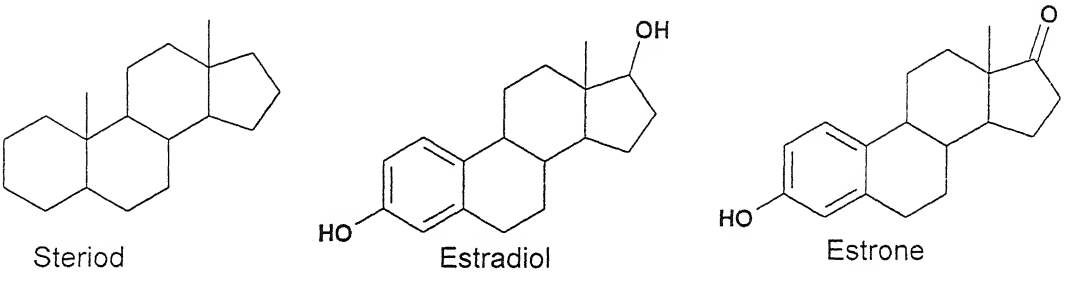


Figure:1

The most prevalent forms of human estrogen are estradiol and estrone (**figure 1**). Both are produced and secreted by the ovaries, although estrone is also made in the adrenal glands and other organs.

Estrogens are hormones, which means that they function as signaling molecules. A signaling molecule exerts its effects by traveling through the bloodstream and interacting with cells in a variety of target tissues. The breast and the uterus, which play central roles in sexual reproduction, are two of the main targets of estrogen. In addition, estrogen molecules act on the brain, bone, liver, and heart.

2.1 ESTROGEN RECEPTORS AND GENE ACTIVATION

Estrogens act on target tissues by binding to parts of cells called *estrogen receptors*. An estrogen receptor is a protein molecule found inside those cells that are targets for estrogen action. Estrogen receptors contain a specific site to which only estrogens (or closely related molecules) can bind. The target tissues affected by estrogen molecules all contain estrogen

receptors; other organs and tissues in the body do not. Therefore, when estrogen molecules circulate in the bloodstream and move throughout the body, they exert effects only on cells that contain estrogen receptors. Estrogen receptors normally reside in the cell's nucleus, along with DNA molecules. In the absence of estrogen molecules, these estrogen receptors are inactive and have no influence on DNA (which contains the cell's genes). But when an estrogen molecule enters a cell and passes into the nucleus, the estrogen binds to its receptor, thereby causing the shape of the receptor to change (**figure 2**). This estrogen-receptor complex then binds to specific DNA sites, called estrogen response elements, which are located near genes that are controlled by estrogen. After it has become attached to estrogen response elements in DNA, this estrogen-receptor complex binds to coactivator proteins and more nearby genes become active. The active genes produce molecules of messenger RNA, which guide the synthesis of specific proteins. These proteins can then influence cell behavior in different ways, depending on the cell type involved.

The estrogen receptor is a ligand-modulated transcription factor that regulates the activity of certain genes[138]. A member of the nuclear hormone receptor gene superfamily, ER has a multidomain structure, with two conserved domains that are responsible for DNA binding on one hand, and ligand binding, dimerization, and transcriptional activation on the other.[138] The binding of ligands to the hormone-binding domain of ER stablizes the interaction of the receptor with target sequences in the regulatory region of these genes. This binding may be either directly to specific DNA enhancer sequences or in some cases to AP1 enhancers through the AP1 transcription factors Fos and Jun. The activation or repression of these genes by the ligand receptors complex is then mediated by the recruitment of ER of a variety of coregulatory proteins that interact with components of the basal transcriptional complex and have enzymatic activity that alters the architecture of chromatin[139].

For a compound to have an estrogenic or antiestrogenic activity through this complex pathway however, it must first bind to the estrogen receptor protein. There are two estrogen receptors, designated as estrogen receptor α and β (ER α and ERβ)[140,142]. These receptor subtypes are related in both structure and function but they have different tissue distributions[143,144] and somewhat different amino acid sequences in their ligand binding domain[145]. Thus these two ER subtypes have somewhat different ligand-binding characteristics and gene-activating activity[146], although much less is known ERβ than ERα.

Because estrogens can act through different ER subtypes, and the ligand-ER complex can utilize different genes, a variety of different response elements, and in different cells, varied levels of different coregulatory proteins, it is not surprising that the pharmacology of estrogenic compounds is complex[147].

Transcriptional activation is mediated by two different activation functions, one of which is controlled by ligand binding (AF-2). A family of proteins called transcriptional coactivators interact with agonist-bound receptors to mediate transcription. This interaction occurs through one or more Nuclear Receptor interaction regions, or NR boxes, which contain the conserved LXXLL sequence motif. The p160 family of coactivators contain multiple NR boxes that recognize different NRs with varying affinities. For the steroid receptors, function is also controlled by the binding of a large chaperone complex that includes Hsp-90 to the ligand binding domain (LBD). Formation of this complex is apparently required for maintaining the receptor in a ligand binding-competent state. Upon ligand binding, the chaperones dissociate, allowing the receptor to bind DNA and regulate transcription.

The ERa LBD can bind to pure agonists such as the endogenous estrogen, 17b-estradiol (E2) or the synthetic estrogen diethylstilbesterol

(DES), pure antagonists such as ICI-164,384. Other compounds such as tamoxifen and raloxifene (RAL) act as antagonists in particular tissue and promoter contexts. This agonist/antagonist behavior is clinically important for treating cancer. Recently, breast cancer prevention trials with tamoxifen showed a 45% reduction in breast cancer incidence and a decreased occurrence of bone fractures. However, a significant increase in incidence of endometrial cancer was also reported.

We are particularly interested in understanding how ligands modulate transcriptional activity and the role of molecular chaperones in this process. This is a critical step in the rational optimization of compounds for the successful treatment and prevention of breast and other cancers. Furthermore, lessons learned from studies on the ER should be applicable to the broad family of NRs. As a first step in this process, we have recently solved the structures of the ERa LBD bound to either the antagonist 4-hydroxy tamoxifen (OHT; the active tamoxifen metabolite) or to the synthetic steroid DES and a peptide from the GRIP1 coactivator NR Box 2. This work is in collaboration with Dr. Geoffrey Greene, U. Chicago. These structures have been quite informative about the coupling between ligand binding and the functional state of AF-2.

Previous work had indicated that although E2 and RAL bind at the same site within the core of the ERa LBD, each of these ligands induces a different conformation of the last helix in the LBD, helix 12. With agonist bound, helix 12 packs against helices 3,5/6 and 11; by contrast with antagonist bound the position of helix 12 is quite different - now occupying a hydrophobic groove constructed from helices 3 and 5. Mutagenesis has shown that residues in this cleft and on helix 12 form part of the AF-2 recognition surface. In our DES complex, the NR box peptide is bound in an a-helical conformation by the hydrophobic groove formed from helices 3,4,5, and 12. In the OHT complex, instead of forming part of AF-2, helix 12 binds to, and occludes, the coactivator recognition box using an

LXXML motif to mimic the LXXLL from the NR box. The positioning of helix 12 is directed by effects on the secondary and tertiary structure of the LBD programmed by ligand binding. Agonists stabilize secondary structural elements, extending the lengths of helices 3,8 and 11. This then shortens the loop between helices 11 and 12, which in turn does not allow helix 12 to fit in the hydrophobic coactivator pocket. The precise geometry of the ligand and its interactions with different regions of LBD lead to a differential stabilization.

2.2 ESTROGEN-INDUCED STIMULATION OF CELL PROLIFERATION

In some target tissues, the main effect of estrogen is to cause cells to grow and divide, a process called *cell proliferation.*

In breast tissue, for example, estrogen triggers the proliferation of cells lining the milk glands, thereby preparing the breast to produce milk if the woman should become pregnant.

Estrogen also promotes proliferation of the cells that form the inner lining, or *endometrium*, of the uterus, thereby preparing the uterus for possible implantation of an embryo (**figure 3**). During a normal menstrual cycle, estrogen levels fall dramatically at the end of each cycle if pregnancy does not occur. As a result, the endometrium disintegrates and is shed from the uterus and vagina in a bleeding process called *menstruation.*

2.3 ESTROGEN: BENEFICIAL AND HARMFUL EFFECTS

Paradoxically, estrogen can be both a beneficial and a harmful molecule **(figure-4).**

The main beneficial effects of estrogen include its roles in

1. Programming the breast and uterus for sexual reproduction,

2. Controlling cholesterol production in ways that limit the buildup of plaque in the coronary arteries, and

3. Preserving bone strength by helping to maintain the proper balance between bone buildup and breakdown.

The decreased production of ovarian steroids which occurs after the climacteric has been linked to a number of postmenopausal pathologies. These include osteoporosis, coronary heart disease, hot flushes and vaginal dyspareunia[148,149]. Indeed, the clinical use of long term estrogen-based hormone replacement therapy (ERT) in post-menopausal women has proven to be highly effective method for reducing the risks associated with these degenerative diseases. However inspite of the fact that the positive effects of such long-term ERT are increasingly accepted[150], the benefits are achieved at the expense of a number of negative side effects, including uterine bleeding, endometrial hyperplasia, endometrial cancer, and an increased risk of developing breast cancer. The uterine side effects, however, may be reduced by co-treatment with progesterone therapy. These negative side effects, in turn, frequently lead to a reduced patient compliance and reluctance to accept ERT as a treatment form[151].

2.4 CANCER ARISES FROM DNA MUTATIONS IN CELLS

Cancer is caused by DNA damage (i.e., mutations) in genes that regulate cell growth and division.

Some mutations are inherited, while others are caused by exposure to radiation or to mutation-inducing chemicals such as those found in cigarette smoke. Mutations also can occur spontaneously as a result of mistakes that are made when a cell duplicates its DNA molecules prior to cell division **(figure-5)**.

When cells acquire mutations in specific genes that control proliferation, such as proto-oncogenes or tumor suppressor genes, these changes are copied with each new generation of cells. Later, more mutations in these altered cells can lead to uncontrolled proliferation and the onset of cancer.

2.5 ESTROGEN-INDUCED PROLIFERATION OF EXISTING MUTANT CELLS

Although estrogen does not appear to directly cause the DNA mutations that trigger the development of human cancer, estrogen does stimulate cell proliferation **(figure-6)**.

Therefore, if one or more breast cells already possesses a DNA mutation that increases the risk of developing cancer, these cells will proliferate (along with normal breast cells) in response to estrogen stimulation. The result will be an increase in the total number of mutant cells, any of which might thereafter acquire the additional mutations that lead to uncontrolled proliferation and the onset of cancer.

In other words, estrogen-induced cell production leads to an increase in the total number of mutant cells that exist. These cells are at increased risk of becoming cancerous, so the chances that cancer may actually develop are increased.

2.6 ESTROGEN-INDUCED PROLIFERATION AND SPONTANEOUS NEW MUTATIONS

Even in women who do not have any mutant breast cells, estrogen-induced proliferation of normal breast cells may still increase the risk of developing cancer.

The reason involves DNA. A cell must duplicate its DNA molecules prior to each cell division, thereby ensuring that the two new cells resulting from the process of cell division each receive one complete set of DNA

molecules (**figure-7**). But the process of DNA duplication occasionally makes mistakes, so the resulting DNA copies may contain a small number of errors (i.e., mutations). If one of these spontaneous mutations occurs in a gene that controls cell growth and division, it could lead to the development of cancer.

Proliferation of normal cells from exposure to estrogen creates a vulnerability to spontaneous mutations, some of which might represent a first step on the pathway to cancer.

2.7 ESTROGEN AND BREAST CANCER

During each menstrual cycle, estrogen normally triggers the proliferation of cells that form the inner lining of the milk glands in the breast.

If pregnancy does not occur, estrogen levels fall dramatically at the end of each monthly menstrual cycle. In the absence of high estrogen levels, those milk gland cells that have proliferated in any given month will deteriorate and die, followed by a similar cycle of cell proliferation and cell death the following month. For the average woman, this means hundreds of cycles of breast cell division and cell death repeated over a span of roughly 40 years, from puberty to menopause.

2.8 ESTROGEN AND UTERINE CANCER

In the uterus, estrogen triggers the proliferation of endometrial lining cells during each month of the menstrual cycle, followed by death of these cells during menstruation. Over a span of 40 years, from puberty to menopause, hundreds of cycles of cell division and cell death will occur.

These repeated cycles of estrogen-induced cell division tend to increase the risk of developing cancer in the same two ways as in the breast: Estrogen can stimulate the division of uterine cells that already have

DNA mutations, and it also increases the chances of developing new, spontaneous mutations when estrogen stimulates cell proliferation. Whether the mutations are inherited or spontaneous, estrogen-driven proliferation will increase the number of these altered cells that can ultimately lead to the development of uterine cancer

2.9 ANTIESTROGENS

Since estrogen can promote the development of cancer in the breast and uterus, it seems logical to postulate that substances that block the action of estrogen might be helpful in preventing or treating these two types of cancer **(figure-8)**.

This rationale has led scientists to work on the development of "antiestrogen" drugs that can block the action of estrogens and thereby interfere with, or even prevent, the proliferation of breast and uterine cancer cells. Antiestrogens work by binding to estrogen receptors, blocking estrogen from binding to these receptors. This also blocks estrogen from activating genes for specific growth-promoting proteins.

2.10 SELECTIVE ESTROGEN RECEPTOR MODULATORS (SERMs)

In working on the development of antiestrogens, scientists have made a somewhat surprising discovery. Some drugs that block the action of estrogen in certain tissues actually can mimic the action of estrogen in other tissues.

Such selectivity is made possible by the fact that the estrogen receptors of different target tissues vary in chemical structure. These differences allow estrogen-like drugs to interact in different ways with the estrogen receptors of different tissues. Such drugs are called ***selective estrogen receptor modulators*, or SERMs**, because they selectively stimulate or inhibit the estrogen receptors of different target tissues. For example, a SERM might inhibit the estrogen receptor found in breast cells

but activate the estrogen receptor present in uterine endometrial cells. A SERM of this type would inhibit cell proliferation in breast cells, but stimulate the proliferation of uterine endometrial cells.

In brief these compounds that have a mixed endocrine profile that affords agonistic or antagonistic activity in a tissue specific manner hold the promise of a safer alternative to estrogen[152,153]. Some members of the SERM class were initially called **"antiestrogens"** because for their high affinity binding to estrogen receptors (ERs) and ability to counteract estrogen action. However, this nomenclature has proved inadequate to fully describe the actions of these agents. The most commonly studied SERMs are tamoxifen, raloxifene, droloxifine, tormifene, idoxifene, lower meloxifene, CP-366156, EM-800, GW-5638 and LY 353381 **(figure-9)**. All these SERMS have tissue selective action. They are now being used for conditions associated with aging, hormone responsive cancer, osteoporosis, cardiovascular diseases and serve lipid lowering. SERMs are therefore currently the archtypes for a rich category of drug therapies based on a single molecular target, the ER.

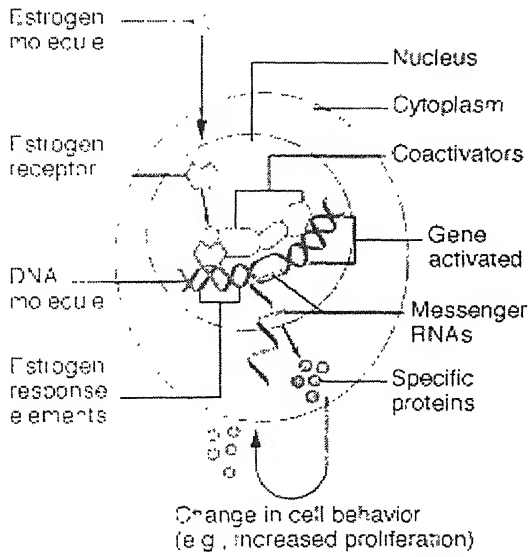


Figure 2

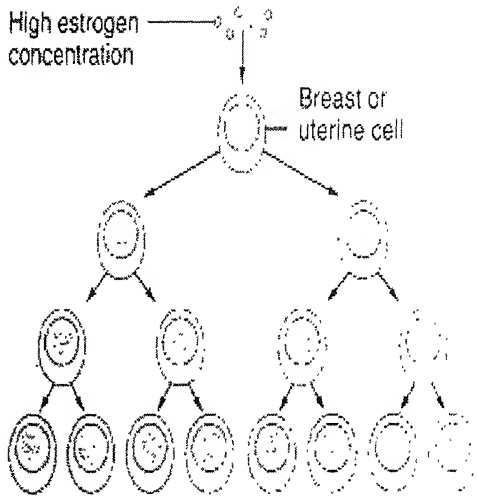


Figure 3

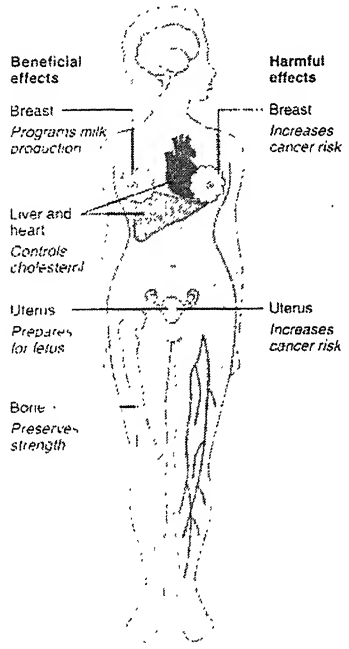


Figure 4

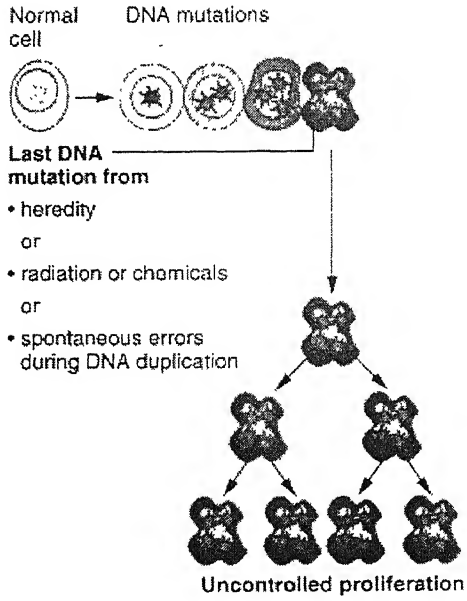


Figure 5

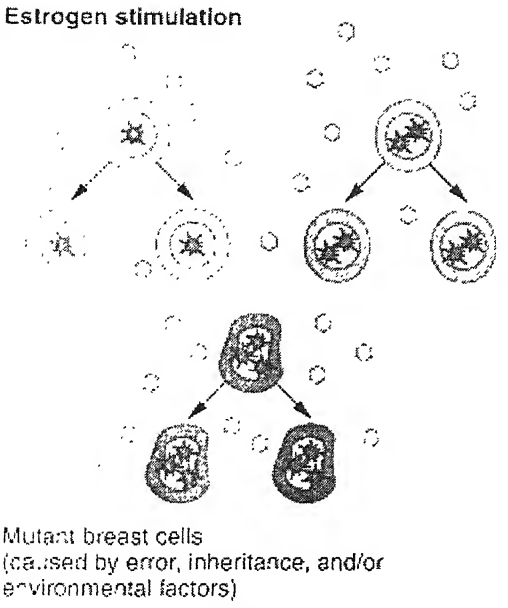


Figure 6

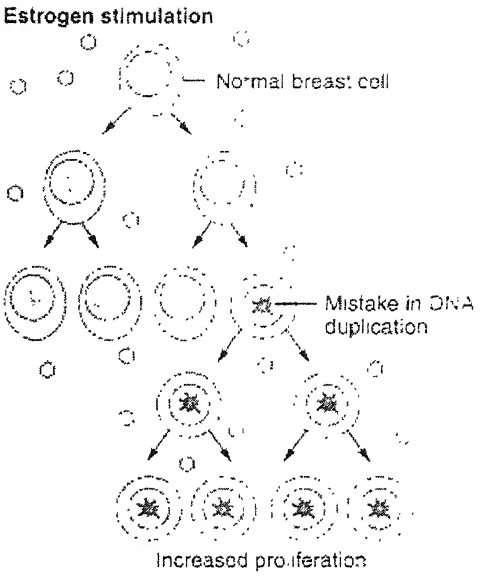


Figure 7

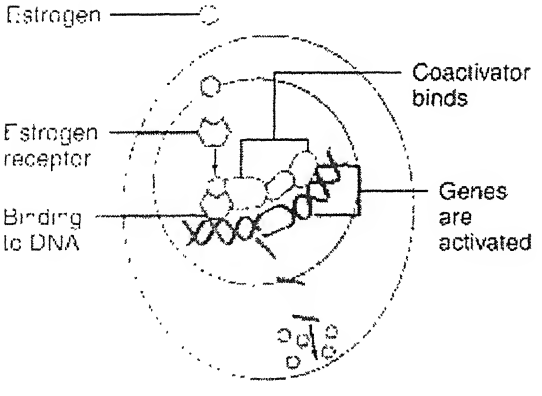


Figure 8A

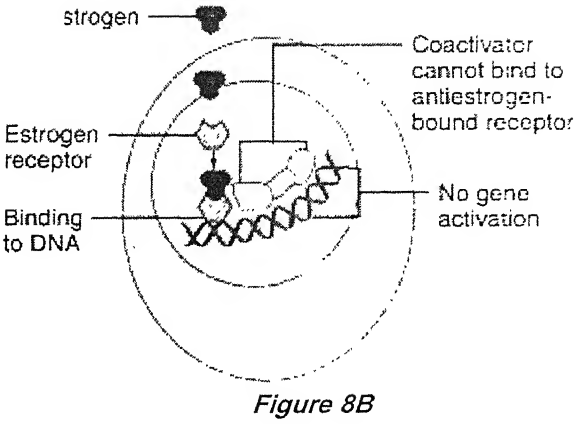
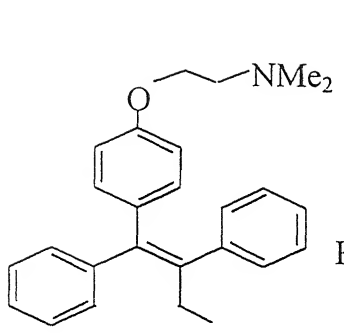
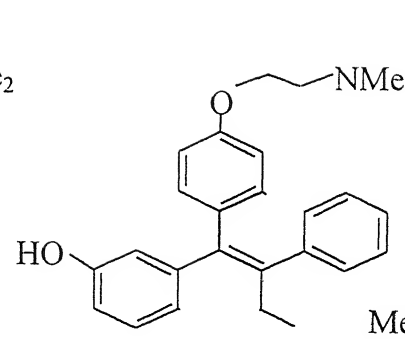


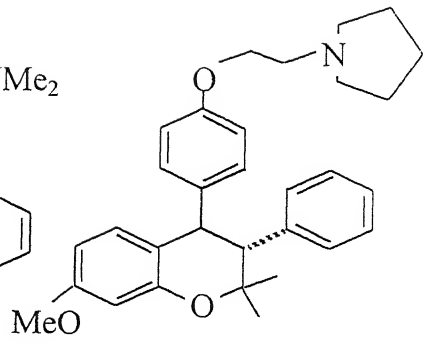
Figure 8B



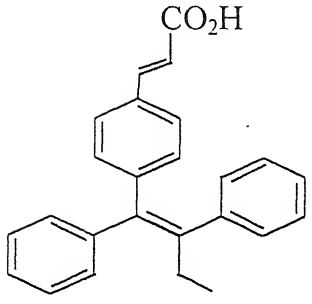
Tamoxifen



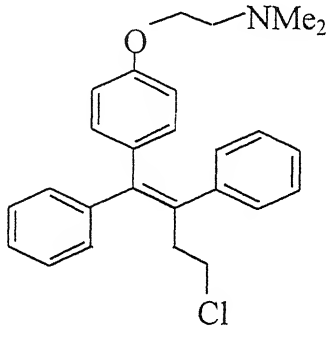
Droloxifene



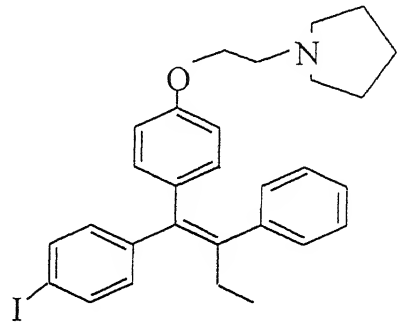
Levormeloxifene



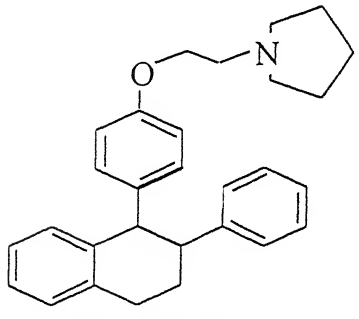
GW5638



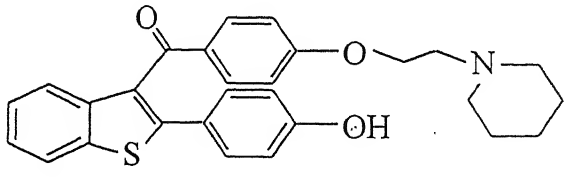
Toremifene



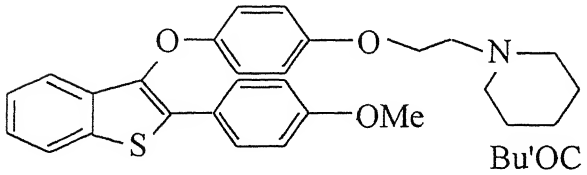
Idoxifene



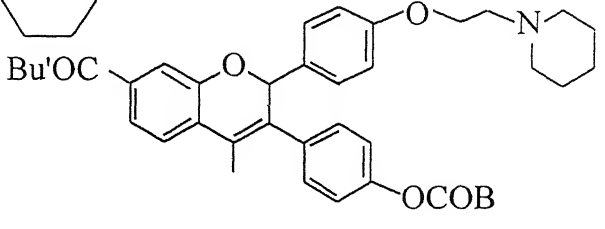
CP - 336156



Raloxifene



LY353381



EM80

Figure-9 : Tissue Selective Estrogenic Compounds.

2.10.1 TAMOXIFEN AND CANCER

The first SERM to be investigated extensively for its anticancer properties is a drug called tamoxifen.

Tamoxifen blocks the action of estrogen in breast tissue. Tamoxifen *exerts* this antiestrogenic effect by binding to the estrogen receptors of breast cells, thereby preventing estrogen molecules from binding to these receptors. But unlike the normal situation, when estrogen binds to its receptor, the binding of tamoxifen to the receptor does not cause the receptor molecule to acquire the changed shape that allows it to bind to coactivators. As a result, the genes that stimulate cell proliferation cannot be activated.

By interfering with estrogen receptors in this way, tamoxifen blocks the ability of estrogen to stimulate the proliferation of breast cells.

2.10.2 TAMOXIFEN AND BREAST CANCER TREATMENT

In women who have breast cancer, proliferation of the breast cancer cells is often driven by estrogen, just as in the case of normal breast cells.

Since tamoxifen can block the effects of estrogen on breast cells, scientists predicted that breast cancer could be treated by using tamoxifen to interfere with estrogen-induced cell proliferation. Based on encouraging results obtained in experimental trials, tamoxifen was first approved for such use in breast cancer treatment in the 1970s.

The first step in treating women with breast cancer is to surgically remove the cancer from the breast. It is difficult to be certain that every cancer cell has been removed at the time of surgery because some breast cancer cells could have spread to surrounding tissues or other organs prior to the operation. Therefore, women often receive some type of

treatment after surgery (*adjuvant therapy*) to prevent the growth of any cancer cells that might remain in the body. Studies show that when tamoxifen is used for this purpose, the risk of cancer recurrence is reduced

2.10.3 TAMOXIFEN AS A CAUSE OF UTERINE CANCER

Although tamoxifen has been useful both in treating breast cancer patients and in decreasing the risk of getting breast cancer in women at high risk, it also has some serious side effects.

These side effects arise from the fact that while tamoxifen acts as an antiestrogen that blocks the effects of estrogen on breast cells, it mimics the actions of estrogen in other tissues such as the uterus. Its estrogen-like effects on the uterus stimulate proliferation of the uterine endometrium and increase the risk of uterine cancer.

2.10.4 SEARCH FOR THE PERFECT SERM

Because of the potential cancer and cardiovascular risks inherent in hormone pills containing estrogen and progesterone, scientists are working on the development of SERMs for postmenopausal women that can mimic the beneficial effects of estrogen without exerting any of its harmful effects.

The ideal drug, of course, would be a SERM exhibiting the positive effects of estrogen on bones, heart, and blood vessels, without exhibiting the potentially harmful effects of estrogen on the breast and uterus.

2.10.5 RALOXIFENE AND THE PREVENTION OF OSTEOPOROSIS

One SERM that may exhibit some of these properties is raloxifene, a drug approved by the FDA in 1997 for preventing osteoporosis in postmenopausal women.

Raloxifene appears to function like estrogen in bone, acting to maintain bone strength and increase bone density. In addition, raloxifene also resembles estrogen in its ability to lower LDL cholesterol levels, thereby decreasing the risk of heart disease.

Although information on the long-term risks and benefits of raloxifene is limited compared to tamoxifen, preliminary evidence suggests that raloxifene may exert these beneficial effects on bones, heart, and blood vessels without increasing a woman's risk of developing cancer

2.10.6 RALOXIFENE AND THE POSSIBLE PREVENTION OF CANCER

Preliminary evidence suggests that raloxifene may actually turn out to be helpful in preventing cancer.

In animal studies, raloxifene has already been shown to reduce the incidence of both breast and uterine cancer. And in preliminary human trials, raloxifene has been found to reduce the risk of breast cancer without the unwanted stimulation of uterine cell division that is exhibited by tamoxifen.

As a result of these preliminary findings, the National Cancer Institute is sponsoring a human clinical study to directly compare the effects of tamoxifen and raloxifene in postmenopausal women. The trial, named STAR (Study of Tamoxifen and Raloxifene), was begun in 1999 and will follow more than 20,000 women for a period of 5 to 10 years.

Even if the STAR trial confirms the effectiveness of raloxifene in reducing the risk of breast and uterine cancer, raloxifene is still not the perfect drug. It does not reduce the frequency of hot flashes associated with menopause and, like estrogen, it increases the risk of blood clots. Just as tamoxifen was an important milestone, if a single SERM like raloxifene is found to protect women against osteoporosis, heart disease, breast cancer, and uterine cancer, it will represent an important milestone in women's health.

The present thesis deals with the applications of QSAR techniques to the following types of non-steroidal SERMs except estradiol-16α-carboxylic acid ester which ins an estradiol derivatiove :

1. **Pyrazole ligands**
2. **Estradiol-16α-carboxylic Acid Esters**
3. **Diaryl-Dialkyl-Substituted Pyrazoles.**
4. **Anti-Estrogen Binding site (AEBS)**
5. **2-Amino-4, 6-diarylpyridines**
6. **cis-3, 4-Diaryl-hydroxy chromanes.**

All the above mentioned ligands are non-steroidal in nature except estradiol-16α-carboxylic acid esters which belong to the class of steroidal estrogens. All these ligands have high affinity and selectivity for estrogen receptor ligands on whom QSAR has been done is determined by different competitive binding affinity assay methods and have been divided into two scales: **"Relative Binding Affinity (RBA) scale** and **Estrogen receptor (ER) binding (IC_{50}).**

It probably does not make a great difference what species and target tissue is used as the source of the estrogen receptor for these binding studies because their is little evidence for species difference in structure affinity relationships, and in most of the target tissue used, the ERα subtype will predominate[157].

All the physicochemical parameters used in the present were automatically loaded from **DRAGON Software** developed by Todeschini, R. and Consonni, V. et. al[158] and the QSAR regression analysis were executed on **Compaq PC using SPSS Software version 6.0.1**.

CHAPTER-3

2D-QSAR RESULTS AND DISCUSSION

This chapter has been divided into two sections according to the binding affinities of estrogen receptor ligands. In the first section, all the binding affinity values were placed on a common "relative binding affinity" (RBA) scale. The RBA values determinations have been done by different competitive binding affinity assay methods and using different receptor preparations. Values on this scale were calculated as a percent from the ratio of IC_{50} values of test compounds to that of estradiol to displace 50% of [^{3}H] estradiol from estrogen receptor preparations (generally uterine cytosol fractions which are largely estrogen receptor-α). Thus on the scale, estradiol by definition has a value of 100, with lower affinity ligands having lower values and higher affinity ligands, higher values.

In the second section the binding affinity values are expressed in IC_{50} forms either extracted from MCF-7 cells lysate by competing with [^{3}H] 17β-estradiol or in estrogen receptor rich cytosol derived from rabbit uterine tissue in a dextran charcoal (DCC) assay[159,160].

SECTION-A

3.1.1 QSAR STUDIES ON PYRAZOLE LIGANDS-ESTROGEN RECEPTOR-α-SELECTIVE AGONISTS[161]

Compounds in this series include various tetrasubstituted pyrazoles as high affinity ligands for the estrogen receptor. Stauffer et al have reported the RBA value in purified full length human ERα and ERβ using competitive radiometric binding assay[158-160].

In this section QSAR have been performed on tetrasubstituted pyrozales on both ERα and ERβ subtypes using hydrophobic (M Log P), steric (MR and V_W) and electronic (Xeq) descriptors which have been described in **Chapter 1**. Regression analysis has been used to correlate the various descriptors with relative binding affinity (RBA) values. **Table 3.1.1.A** contains a set of 15 compounds showing log RBA values of ERα and ERβ subtypes along with the data for regression analysis. **Table 3.1.1.B and Table 3.1.1.C** represents the correlation matrix of ERα and ERβ subtypes with the descriptors used in the present work.

3.1.1.A QSAR STUDY OF TETRA SUBSTITUTED PYRAZOLE LIGNADS BINDING TO ERα SUBTYPES.

The correlation of M Log P, MR, V_W and Xeq. with RBA ERα gave the following simple regression (monoparameteric) equation.

log RBA = 0.372 (± 0.282) M Log P + 0.302 → (1)

n = 16 r^2 = 0.215 SE = 0.512 $F_{(1, 14)}$ = 0.957

log RBA = 0.661 (± 0.314) MR + 0.817 → (2)

n = 16 r^2 = 0.183 SE = 0.625 $F_{(1, 14)}$ = 0.440

log RBA = - 0.418 (± 0.612) V_W + 0.761 → (3)

n = 16 r^2 = 0.017 SE = 0.688 $F_{(1, 14)}$ = 0.393

log RBA = - 0.768 (± 0.427) Xeq + 0.378 → (4)

n = 16 r^2 = 0.369 SE = 0.509 $F_{(1, 14)}$ = 1.630

The monoparameteric equations here shows very poor correlations. **Table 3.1.1.B** shows that neither M Log P correlates will with MR and Xeq. nor MR correlates with Xeq., so these parameters can be taken together. The correlation of M Log P and MR, M Log P and Xeq. and MR and Xeq. is given as follows:

$$\log RBA = -0.810\ (\pm\ 0.513)\ M\ Log\ P + 0.468\ (\pm\ 0.211)\ MR + 0.326 \rightarrow (5)$$

$$n = 16\ \ r^2 = 0.710\ \ SE = 0.443\ \ F_{(2,\ 13)} = 6.94$$

$$\log RBA = -0.783\ (\pm\ 0.443)\ M\ Log\ P + 0.660\ (\pm\ 0.431)\ Xeq. + 0.684 \rightarrow (6)$$

$$n = 16\ \ r^2 = 0.768\ \ SE = 0.410\ \ F_{(2,\ 13)} = 8.39$$

$$\log RBA = 0.572\ (\pm\ 0.418)\ MR - 0.54\ (\pm\ 0.529)\ Xeq. + 0.376 \rightarrow (7)$$

$$n = 16\ \ r^2 = 0.472\ \ SE = 0.510\ \ F_{(2,\ 13)} = 1.96$$

Equations (5) and (6) show significant improvement in correlation level. They account for 71% and 77% variance ratio (R = 0.84 and R = 0.88) and F-value is also significant at 95% confidence interval. Equation (7) however cannot be accepted from statistical point of view. The error terms in this equation are quite high and the F-value is not significant at 95% confidence interval. Now in an attempt to improve the degree of correlation of equations (5) and (6), the introduction of an indicator parameter (Ind) for substituents having isobutyl group gave fairly good improvement in correlations:

$$\log RBA = -0.932\ (\pm\ 0.476)\ M\ Log\ P + 0.511\ (\pm\ 0.310)\ MR + 0.174\ (\pm\ 0.043)\ Ind + 0.528 \rightarrow (9)$$

$$n = 16\ \ r^2 = 0.796\ \ SE = 0.462\ \ F_{(3,\ 12)} = 7.886$$

$$\log RBA = -0.872\ (\pm\ 0.342)\ M\ Log\ P - 0.83\ (0.410)\ Xeq + 0.233\ (\pm\ 0.008)\ Ind + 0.382 \rightarrow (10)$$

$$n = 16\ \ r^2 = 0.865\ \ SE = 0.318\ \ F_{(3,\ 12)} = 11.204$$

Equation (10) accounts for 87% variance ratio (R = 0.93) and F-value is significant at 99% confidence level.

3.1.1.BQSAR STUDY OF TETRASUBSTITUTED PYRAZOLE LIGANDS BINDING TO ERβ SUBTYPE.

The correlation of M Log P, MR, Vw and Xeq with log RBA gave simple linear regression equation as:

$$\log RBA = -0.424 (\pm 0.317) \text{ M Log P} + 0.243 \quad \rightarrow (11)$$
$$n = 16 \quad r^2 = 0.281 \quad SE = 0.510 \quad F_{(1, 14)} = 0.865$$
$$\log RBA = 0.48 (\pm 0.382) \text{ MR} + 0.776 \quad \rightarrow (12)$$
$$n = 16 \quad r^2 = 0.136 \quad SE = 0.622 \quad F_{(1, 14)} = 0.648$$
$$\log RBA = 0.634 (\pm 0.521) \text{ Vw} - 0.938 \quad \rightarrow (13)$$
$$n = 16 \quad r^2 = 0.004 \quad SE = 0.821 \quad F_{(1, 14)} = 0.402$$
$$\log RBA = -0.245 (\pm 0.118) \text{ Xeq} + 0.239 \quad \rightarrow (14)$$
$$n = 16 \quad r^2 = 0.189 \quad SE = 0.712 \quad F_{(1, 14)} = 0.461$$

The regression analysis shows very poor correlation with all the four parameters (equations 11, 12, 13 and 14). From **Table 3.1.1.C.** good autocorrelation is observed between M Log P and Vw, M Log P and Xeq, MR and Vw and Vw and Xeq so these descriptors cannot be taken together in multiple regression analysis. However M Log P and MR and MR and Xeq do not show any autocorrelation so these parameters have been taken together and MRA performed shows significant improvement in correlation.

$$\log RBA = -0.830 (\pm 0.511) \text{ M Log P} + 0.66 (\pm 0.511) \text{ MR} + 0.268 \quad \rightarrow (16)$$
$$n = 16 \quad r^2 = 0.588 \quad SE = 0.416 \quad F_{(2, 13)} = 0.87$$
$$\log RBA = 0.512 (\pm 0.396) \text{ MR} - 0.410 (\pm 0.245) \text{ Xeq} + 0.211 \quad \rightarrow (17)$$
$$n = 16 \quad r^2 = 0.402 \quad SE = 0.511 \quad F_{(2, 13)} = 4.662$$

Introducing the same indicator parameter, Ind, that was used in ERα subtype in equations (16) and (17) statistically significant correlations were obtained.

$$\log RBA = -0.741\ (\pm 0.45)\ M\ Log\ P + 0.762\ (\pm 0.410)\ MR + 0.41\ (\pm 0.23)\ Ind + 0.723 \quad \rightarrow (18)$$

$$n = 16\ \ r^2 = 0.786\ \ SE = 0.376\ \ F_{(3,\ 12)} = 9.832$$

$$\log RBA = 0.632\ (\pm\ 0.307)\ MR - 0.397\ (\pm\ 0.197)\ Xeq + 0.33\ (\pm 0.201)\ Ind + 0.510 \quad \rightarrow (19)$$

$$n = 16\ \ r^2 = 0.713\ \ SE = 0.504\ \ F_{(3,\ 12)} = 7.314$$

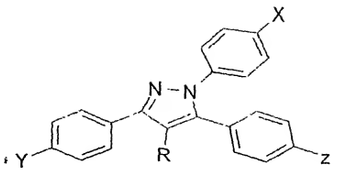
However, with a hope of obtaining still better results, compound (16) having the highest residual value from equation (18) was taken as an outlier and the following regression equation has been obtained:

$$\log RBA = -0.543\ (\pm 0.327)\ M\ Log\ P + 0.821\ (\pm 0.372)\ MR + 0.412\ (\pm 0.212)\ Ind + 0.417 \quad \rightarrow (20)$$

$$n = 15\ \ r^2 = 0.845\ \ SE = 0.301\ \ F_{(3,\ 12)} = 11.46$$

Equation (20) accounts for 85% variance ratio (R = 0.92) and F-value is significant at 99% confidence interval.

TABLE-3.1.1.A : PHYSICO-CHEMICAL DATA FOR PYRAZOLE LIGANDS



S. No.	COMPD	SUBSTITUENTS				DESCRIPTORS				OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResRBAERα	OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResRBAER β
		R	X	Y	Z	MLogP	MR	Xeq	Vw	RBAERα	LogRBAERα	PreRBAERα Eq. (10)		RBAERβ	LogRBAERβ	PreRBAERβ Eq. (20)	
1.	4a	Me	H	OH	OH	2.76	0.17	26.99	24.41	0.76	-0.12	-0.12	-	0.28	-0.55	-0.66	0.11
2.	4b	Et	H	OH	OH	2.61	0.15	31.32	27.93	31	+1.49	1.86	-0.37	1.1	+0.04	+0.17	-0.13
3.	6A	i-Pr	H	OH	OH	3.79	0.17	26.66	24.90	5.6	+0.75	0.84	-0.09	0.86	-0.07	-0.22	0.15
4.	4c	n-Pr	H	OH	OH	3.79	0.17	26.66	24.90	16.8	+1.23	0.84	0.46	0.52	-0.28	0.22	-0.06
5.	4d	i-Bu	H	OH	OH	3 77	0.17	26.66	25.41	56	+1.75	1.21	0.54	1.4	+0.15	0.27	-0.12
6.	4e	n-Bu	H	OH	OH	2.98	0.17	26.66	25.41	8.7	+0.93	0.61	0.33	0.47	-0 33	-0.28	-0.05
7.	4f	Et	OH	OH	OH	2.91	0.23	27.99	25.41	36	+1.56	1.32	0.24	0.15	-0.82	-0.7	-0.12
8.	4g	n-Pr	OH	OH	OH	2.91	0.23	27.99	25 41	49	+1.69	1.32	0.37	0.12	-0.92	-0.7	-0.22
9.	4h	i-Bu	OH	H	OH	2.91	0 23	27.99	26.41	75	+1 88	1.62	0 26	0.89	-0.05	-0.17	0.12
10.	4i	n-Bu	OH	OH	OH	3.08	0 26	27.99	26.41	14	+1.15	1.49	-0.34	0.18	-0.74	-1.07	0 33
11.	7a	Et	OH	H	H	3.12	0.23	28 99	27.41	3 1	+0.49	0.63	-0 11	1.5	+0.18	-0.38	-0.2
12.	7b	Et	H	OH	H	3.33	0.22	28.99	27.41	2.6	+0 41	-0.73	-0.32	0.61	-0 21	-0.05	-0 16
13.	7c	Et	H	H	OH	3.33	0.22	29.99	25.93	0 04	-1.4	-0.73	-0.67	0 06	-1 22	-0 05	0.21
14.	4b	Et	H	OH	OH	2.12	0 23	29.99	26.93	31	+1.49	0.86	0 63	1.1	+0.04	0 12	-0.08
15.	8a	Et	OH	H	OH	2 33	0.22	29 32	27.93	7	+0.85	-0.19	1 04	0 8	-0 1	0 48	0 38
16.	8b	Et	OH	OH	H	2.54	0.21	30.32	27.93	8.9	+0.95	1.03	-0.08	0 32	-0 49	-1.03	0 54

[74]

CORRELATION MATRIX

TABLE: 3.1.1.B

(FOR ERα SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.448	1.000		
Vw	0.516	0.671	1.000	
Xeq	0.146	0.230	0.637	1.000

TABLE: 3.1.1.C

(FOR ERβ SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.203	1.000		
Vw	0.672	0.890	1.000	
Xeq	0.546	0.149	0.771	1.000

3.1.2 QSAR STUDIES ON ESTRADIOL-16α-CARBOXYLIC ACID ESTERS AS LOCALLY ACTIVE ESTROGENS[162]

Compounds in this series include various analogues of estradiaol acting as locally active estrogens without significant systemic action useful for the therapeutic treatment of vaginal dyspareunia of menopause in women for whom systemic estrogens are contraindicated[163]. The RBA values were studied using the classical assay with rat uterine cytosol (ERα). In this section QSAR have been performed on estradiol-16α-carboxylic acid esters. The relative binding affinity data was correlated with hydrophobic (M Log P), steric (MR and Vw) and electronic (Xeq) descriptors.

A complete set of molecular descriptors, namely: M Log P, MR, Vw, Xeq, indicator parameter for a set of 16 compounds along with the log RBA values are recorded in **Table 3.1.2.A** and **Table 3.1.2.B** represents the correlation matrix between the parameters. Correlation matrix indicates that their is very good autocorrelation between M Log P and Vw, MR and Xeq and Vw and Xeq, that is why these parameters cannot be taken together in multiple regression analysis (MRA).

On applying simple linear regression analysis, the correlation of relative binding affinity data with M Log P, MR, Xeq and Vw, gave simple linear equation as:

$$\log RBA = -0.189\ (\pm 0.069)\ M\ Log\ P + 2.155 \quad \rightarrow (1)$$
$$n = 16\ \ r^2 = 0.401\ \ SE = 0.38\ \ F_{(1,\ 14)} = 5.20$$
$$\log RBA = 0.614\ (\pm 0.332)\ MR + 0.487 \quad \rightarrow (2)$$
$$n = 16\ \ r^2 = 0.289\ \ SE = 0.732\ \ F_{(1,\ 14)} = 0.013$$
$$\log RBA = 0.671\ (\pm 0.326)\ Vw + 1.286 \quad \rightarrow (3)$$
$$n = 16\ \ r^2 = 0.321\ \ SE = 0.501\ \ F_{(1,\ 14)} = 1.281$$
$$\log RBA = -0.298\ (\pm 0.836)\ Xeq - 1.120 \quad \rightarrow (4)$$
$$n = 16\ \ r^2 = 0.129\ \ SE = 0.672\ \ F_{(1,\ 14)} = 0.761$$

Equation (1), (2), (3) and (4) show very poor correlation. **Table 3.1.2.B** shows that their is neither any correlation between M Log P and MR nor between M Log P and Xeq, so these parameters can be taken together. The correlation of RBA value with M Log P and MR is given as:

$$\log RBA = -1.272\ (\pm 0.529)\ M\ Log\ P + 0.921\ (\pm 0.637) MR + 1.216 \quad \rightarrow (5)$$

$$n = 16\ \ r^2 = 0.768\ \ SE = 0.314\ \ F_{(2, 13)} = 19.27$$

Equation (5) accounts for 76% variance ratio (R = 0.876) and F-value is significant at 99% confidence interval. Similarly, the correlation of RBA value with M Log P and Xeq gave the following regression equation:

$$\log RBA = -0.922\ (\pm 0.646)\ M\ Log\ P - 0.733\ (\pm 0.396) Xeq + 0.839 \quad \rightarrow (6)$$

$$n = 16\ \ r^2 = 0.614\ \ SE = 0.483\ \ F_{(2, 13)} = 8.10$$

Equation (6) accounts for 61% variance ratio (R = 0.784) and F-value is significant at 95% level. Comparing equation (5) and (6), equation (5) seems more statistically sound. In order to improve the degree of correlation an indicator parameter, Ind has been used. Indicator parameter has been taken for estradiol having neo-pentyl ester, the value for which is taken as unity for the presence of this ester and for all other compounds where this ester is absent, it is taken as zero.

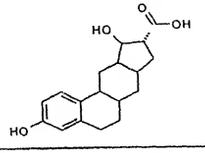
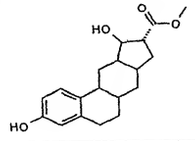
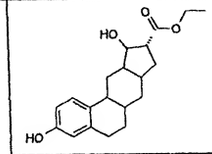
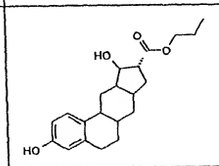
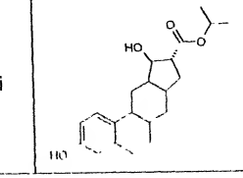
MRA gave the following triparametric model which shows a significant improvement in correlation as follows:

$$\log RBA = 1.272\ (\pm 0.429)\ M\ Log\ P + 0.821\ (\pm 0.437)\ MR + 0.537\ (\pm 0.311)\ Ind + 1.889 \quad \rightarrow (7)$$

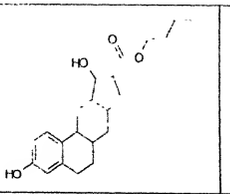
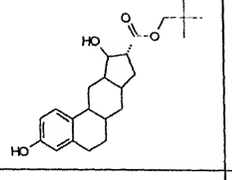
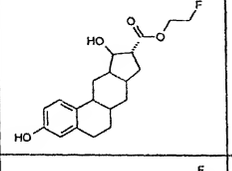
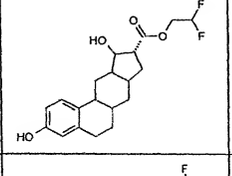
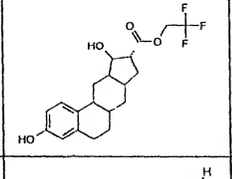
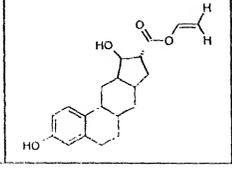
$$n = 16\ \ r^2 = 0.910\ \ SE = 0.226\ \ F_{(3, 10)} = 18.33$$

Equation (7) accounts for 91% variance ratio (R = 0.954) and F-value is significant at 99% confidence interval.

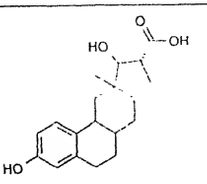
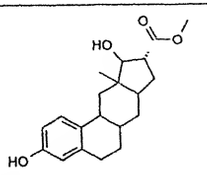
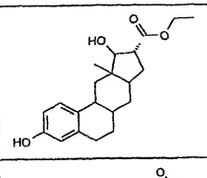
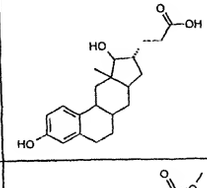
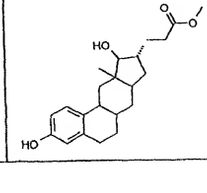
TABLE-3.1.2.A : PHYSICO-CHEMICAL DATA FOR ESTRADIOL-16α-CARBOXYLIC ACID ESTERS

S. No.	COMPD	STRUCTURE	DESCRIPTORS				OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResRBA
			MLogP	MR	Xeq	Vw	ERRBA	ERLogRBA	PreRBA EQ. (7)	
1.	E16-1,2		2.55	65.73	23.32	20.05	0		.48	
2.	E16-1,1		3.34	76.48	29.69	2328	35	+1.54	.48	1.06
3.	E16-1,2		2.93	75.27	27.21	22.85	40	+1.6	1.69	.09
4.	E16-1,3		2.88	64.48	24.32	21.05	34	+1.53	1.16	.37
5.	E16-1,3i		3.09	68.14	25.32	22.05	7	+0.85	1.16	-.81

[76]

6.	E16-1,4		3.31	71.41	26 32	23 05	28	+1 45	1 16	29
7.	E16-1,5neo		3.39	67	26.32	23.05	14	+1.15	1.23	-.08
8.	E16-1,2F1		3.61	70.66	27.32	24.05	35	+1.54	1.63	-.09
9.	E16-1,2F2		2.78	69.39	24.32	21.05	10	+1	.88	.12
10.	E16-1,2F3		3.01	72.67	25.32	22.05	7	+0.85	.88	-.03
11.	E16-1,2vin		3.23	75 29	26.32	23.05	6	+0.78	.48	.3

[77]

12.	E16-2.0		3.23	74 63	26.32	23 05	0		48	
13.	E16-2,1		3.45	77.91	27.32	24.05	5	+0.7	.48	.22
14.	E16-2,2		3.67	80.53	28.32	25.05	5	+0.7	.48	.22
15.	E16-3,0		3.12	74.51	26.78	22.46	0.1	-1	.48	.52
16.	E16-3,1		3.23	76.01	28.24	22.87	1	0	.48	-.48-.48

CORRELATION MATRIX

TABLE: 3.1.2.B

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.262	1.000		
Vw	0.765	0.449	1.000	
Xeq	0.370	0.994	0.839	1.000

3.1.3 QSAR STUDIES ON DIARYL-DIALKYL-SUBSTITUTED PYRAZOLES[164]

Nishiguchi et al[164] reported regioselective synthesis and binding affinity data of diaryl dialkyl-substituted pyrazoles for the estrogen receptor. Compounds of this class have been found to be useful for menopausal hormone replacement and for the prevention and treatment of breast cancer[165]. The binding affinity for each pyrazole phenol for human estrogen receptors (ERα and ERβ) was determined using a competitive radiometric assay; affinities were expressed relative to that of estradiol, to give relative binding affinity (RBA) values[166]. In this discussion QSAR have been performed on a series of substituted pyrazoles using hydrophobic (M Log P), steric (MR and Vw) and electronic (Xeq) descriptors.

Table 3.1.3.A lists RBA, M Log P, MR, Vw and Xeq values for a set of 8 substituted pyrazoles for both ERα and ERβ subtypes used in the present study. **Table 3.1.3.B** and **Table 3.1.3.C** represents the correlation matrices between the parameters for ERα and ERβ subtypes.

3.1.3.AQSAR STUDY OF SUBSTITUTED PYRAZOLE BINDING WITH ERα AND ERβ SUBTYPE

Simple linear regression analysis using these physicochemical parameters independently resulted in the following regression equations.

$$\log RBA = -0.739\ (\pm 0.437)\ M\ Log\ P + 1.288 \quad \rightarrow (1)$$
$$n = 8\ \ r^2 = 0.528\ \ SE = 0.336\ \ F_{(1,6)} = 3.035$$
$$\log RBA = -0.116\ (\pm 0.018)\ MR + 0.972 \quad \rightarrow (2)$$
$$n = 8\ \ r^2 = 0.768\ \ SE = 0.418\ \ F_{(1,6)} = 12.234$$
$$\log RBA = -0.820\ (\pm 0.512)\ Vw + 1.011 \quad \rightarrow (3)$$
$$n = 8\ \ r^2 = 0.048\ \ SE = 0.821\ \ F_{(1,6)} = 0.326$$
$$\log RBA = -1.114\ (\pm 1.702)\ Xeq + 0.327 \quad \rightarrow (4)$$
$$n = 8\ \ r^2 = 0.118\ \ SE = 0.686\ \ F_{(1,6)} = 0.214$$

Initial study has shown that the steric descriptor, M Log P and MR gave statistically significant correlations while the remaining descriptors gave poor correlations. **Table 3.1.3.B** shows their is no autocorrelation between MR and M Log P, so these descriptors can be taken together.

$$\log RBA = -0.84\ (\pm 0.52)\ M\ Log\ P + 1.162\ (\pm 0.612)\ MR + 0.461 \quad \rightarrow (5)$$

$$n = 8 \quad r^2 = 0.848 \quad SE = 0.378 \quad F_{(2,5)} = 18.99$$

Equation (5) shows improvement in r^2 value and F-value is also significant at 99% confidence interval.

However, with a hope of obtaining still better results, an indicator parameter, Ind, for the substituent having C_6H_{11} group was used with MR. As a consequence excellent correlation was observed.

$$\log RBA = -0.887\ (\pm 0.310)\ MR + 1.018\ (\pm 0.602)\ Ind + 0.761 \quad \rightarrow (6)$$

$$n = 8 \quad r^2 = 0.926 \quad SE = 0.284 \quad F_{(2,5)} = 26.33$$

Equation (6) accounts for 93% variance ratio (R = 0.962) and F-value is significant at 99% confidence interval also the error term has become quite low.

3.1.3.B QSAR STUDY OF SUBSTITUTED PYRAZOLES WITH ERβ SUBTYPES

The correlation of RBA for ERβ subtype with hydrophobic, steric and electronic parameters gave the following regression equations.

$$\log RBA = -0.814\ (\pm 0.562)\ M\ Log\ P + 1.011 \quad \rightarrow (7)$$

$$n = 8 \quad r^2 = 0.324 \quad SE = 0.378 \quad F_{(1,6)} = 2.954$$

$$\log RBA = -0.446\ (\pm 0.274)\ MR + 0.837 \quad \rightarrow (8)$$

$$n = 8 \quad r^2 = 0.505 \quad SE = 0.481 \quad F_{(1,6)} = 5.077$$

$$\log RBA = -0.488\ (\pm 0.502)\ Vw + 0.917 \quad \rightarrow (9)$$

$$n = 8\ \ r^2 = 0.113\ \ SE = 0.669\ \ F_{(1,6)} = 0.310$$

$$\log RBA = -0.217\ (\pm 0.228)\ Xeq + 1.119 \quad \rightarrow (10)$$

$$n = 8\ \ r^2 = 0.032\ \ SE = 0.811\ \ F_{(1,6)} = 0.332$$

Equations (9) and (10) show very poor correlations and (7) and (8) are not very statistically significant either. However introduction of square terms of MR improved the correlation level.

Equation (11) shows an abrupt increase in correlation level as compared to equation (8) (R = 0.712 to R = 0.886). This indicates that MR shows parabolic relation with RBA values of ERβ subtypes, Vw and Xeq do not show any significant improvement in correlation level on introduction of square terms. M Log P however shows a very (modulate) slight improvement in correlation level.

$$\log RBA = -0.882\ (\pm 0.344)\ M\ Log\ P + 0.601\ (\pm 0.446)\ (M\ Log\ P)^2 - 1.022 \quad \rightarrow (12)$$

$$n = 8\ \ r^2 = 0.501\ \ SE = 0.473\ \ F_{(2,5)} = 2.392$$

An order to improve the degree of correlation same indicator parameter which was used in ERα subtype was used with MR and the following regression equation was obtained:

$$\log RBA = 0.761\ (\pm 0.542)\ MR + 0.936\ (\pm 0.623)\ Ind + 0.843 \quad \rightarrow (13)$$

$$n = 8\ \ r^2 = 0.721\ \ SE = 0.396\ \ F_{(2,5)} = 9.311$$

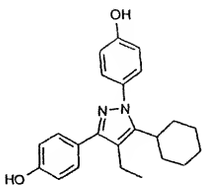
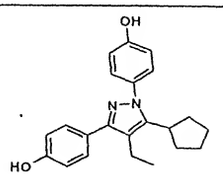
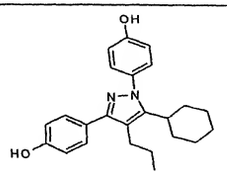
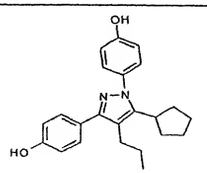
Here in equation (13) as compared to equation (6) only moderate improvement in correlation level is observed. Equation (13) accounts for 72% variance ratio (R = 0.85) and the F-value is significant at 95% confidence level. From **Table 3.3.C** there is no autocorrelation between M Log P and MR, so these descriptors can be taken together. Significant improvement in correlation was observed:

$$\log RBA = - 0.833 (\pm 0.326) M \text{ Log } P + 0.794 (\pm 0.310) MR + 0.831 \rightarrow (14)$$

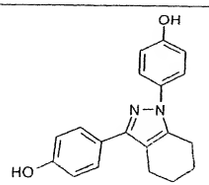
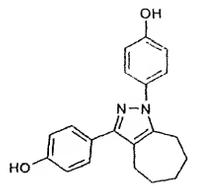
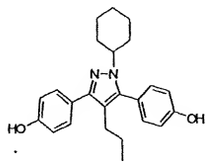
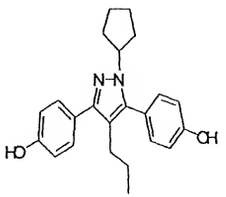
$$n = 8 \quad r^2 = 0.886 \quad SE = 0.297 \quad F_{(2,5)} = 21.033$$

Equation (14) accounts for 89% variance ratio (R = 0.94) and F-value is significant at 99% confidence interval.

TABLE-3.1.3.A : PHYSICOCHEMICAL DATA FOR DIARYL-DIALKYL-SUBSTITUTED PYRAZOLES

S. No.	COMPD	STRUCTURE	DESCRIPTORS				OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResRBAERα	OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResRBAERβ
			MLogP	MR	Xeq	Vw	RBAERα	LogRBAERα	PreRBAERα EQ. (6)		RBAERβ	LogRBAERβ	PreRBAERβ EQ. (14)	
1.	11a		3.2	82.3	25 16	29.99	0.59	-0.23	-0 37	0 14	0.14	-0.85	-1.28	0 43
2.	11b		2.98	80.1	24.16	26.99	2.1	0.32	0.14	0.17	0.23	-0.64	-0.60	-0 03
3.	11c		3.41	85.4	26.16	20.99	1.1	0.04	0.39	-0.35	0.11	-0.96	-0.29	-0.66
4.	11d		3.2	82.8	25.16	29.99	3.6	0.56	0.26	0.29	0 6	-0.22	-0.46	0 24

[84]

5.	15a		2 35	72.3	21.16	23.99	0 56	-0 25	-0 25	0.00	0 04	-1 38	-1 11	0 26
6.	15b		2.58	74.9	22 16	24 99	0.19	-0.72	-0.42	-0 29	0.05	-1 33	-1 33	0 00
7.	15c		3.41	86.5	26.16	28.99	30	1.48	1.79	-0.31	13	1.11	1.51	-0.40
8.	15d		3.2	83.8	25.16	27.99	74	1.87	1.52	0.34	71	1.85	1.17	0.67

CORRELATION MATRIX

TABLE: 3.1.3.B

(FOR ERα SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.112	1.000		
Vw	0.470	0.361	1.000	
Xeq	0.832	0.752	0.681	1.000

TABLE: 3.1.1.C

(FOR ERβ SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.203	1.000		
Vw	0.672	0.890	1.000	
Xeq	0.546	0.149	0.771	1.000

CONCLUSIONS

Thus from the regression analysis data of the three series (3.1.1, 3.1.2, 3.1.3) for relative binding affinity values in the present thesis we see that hydrophobic parameter M Log P and MR plays a crucial role in enhancing the RBA values of the estrogen receptor. Besides electronic parameter, Xeq, also affects the binding affinity to some extent. Vw, however showed poor correlations in all the three series, it means Vw does not plays any significant role in enhancing the binding affinities in the series undertaken in performing QSAR.

The large negative coefficient of M Log P indicates, that highly hydrophobic substituents will lower the binding affinity in this class of ERs however positive coefficient of MR indicates that sterically bulky substituents will raise the binding affinity. Since the relationship with MR is linear, this implies that the receptor has some flexibility at this site. It can very well be inferred that some component of hydrophobicity is embodied in the MR term.

The negative coefficient of electronic parameter, Xeq equation (10) and equation (11) of series 3.1.1, suggests that electron donating groups are favourable for relative binding affinity values. It is of note that the X-ray crystal structure of the receptor shows that there is some very hydrophobic space above the B-ring of the ligand at position 11[167], presumably sufficient to accommodate groups of moderate size without larger substituents would require some movement of ligand or receptor for a complex to form. It can be generalized from the set of congeners studied in this section that the negative hydrophobic and positive steric effect is operative here. However, the fact that correlation taken with indicator parameter here is small and positive, which indicates that substituents chosen as indicator parameter will increase the relative binding affinities markedly.

It is gratifying for estradiol-16α-carboxylic acid esters we see at the 11β-position the same hydrophobic preference (positive correlation with MR). The non-polar neo-pentyl ester (positive indicator parameter) enhances binding affinity. There is no evidence for significant substituent effects at position 16α, although the size range of 16α substituent in this series (3.1.2) is much smaller. Gontcher et al[168] studied the relative binding affinities of estradiol derivatives with multiple substituents at 2- ,4-, 7α-, 11β- and 17α-positions with estrogen receptor in the cytosol of mouse mammary epithelial cells at 0°C. They did 3D-analysis with CoMFA, which revealed the importance of electrostatic and steric fields.

In the case of pyrazole ligands (series 3.1.1 and 3.1.3) the substituents that were on the phenolic ring were minutely examined although the pyrazole nucleus also was not neglected. As expected for aromatic substituents for both ERα and ERβ subtypes, steric bulk interferes with binding. However it seems that electron-donating groups will increase the electron density on the phenyl ring, but will also make the phenolic hydroxyl group less acidic. Thus, the increased affinity of the derivatives with electron donating groups could be due to :

(a) An increased electron transfer interaction between receptor and ligands or

(b) An affect on the electron density on the OH.

From the crystal structure of the estrogen receptor ligand binding domain, both affects are reasonable[167], as the A-ring of estradiol is tightly surrounded by residues and the phenolic hydroxyl group donates one hydrogen bond (which would be weakened by the increased electron density) but accepts two hydrogen bonds.

In brief, from equations (9), (10), (18), (19), (20), of series 3.1.1 (7) of series 3.1.2 and (5), (6), (13), (14) of series 3.1.3 we see that estrogen receptor has a limited tolerance to hydrophobic effects and a vast tolerance to steric effects of the substituents. The receptor site has a hydrophobic binding domain.

SECTION-B

3.2.1 QSAR STUDIES OF NEW LIGANDS FOR THE MICROSOMAL ANTI-ESTROGEN BINDING SITE (AEBS)[169]

Compounds in this series include diphenyl methane derivative, N, N-diethyl-2-[(4-phenyl-methyl)-phenoxy]-ethanamine, HCl (DPPE). These new compounds have no affinity for the estrogen receptor and bind with various affinity to the anti-estrogen binding site (AEBS)[170,171]. These compound have been shown to display potential clinical value as they are cytotoxic against tumor cells[172-174] and display antiviral activities[175]. Poirot et al[167] have evaluated the binding affinity for the ER extracted from MCF-7 cells by competing with [^{3}H] estradiol, and for binding to rat liver microsomal AEBS by competing with [^{3}H] tamoxifen.

In the present study QSAR has been performed on the various ligands for the microsomal on the AEBS. The binding affinities values for AEBS expressed as log Ki was correlated with hydrophobic (M Log P), steric (MR and Vw) and electronic (Xeq) descriptors. **Table 3.2.1.A** lists the pki values along with the physicochemical data for a set of 36 compounds for regression analysis. The correlation between the parameters, which are used in the present work, are given in **Table 3.2.1.B**.

Correlation of hydrophic (M Log P), steric (MR and Vw) and electronic (Xeq) descriptors with the binding affinity gave the following simple linear regression equations:

$$- \log ki = 0.568\ (\pm 0.325)\ M\ Log\ P + 0.461 \qquad \rightarrow (1)$$

$$n = 36\ \ r^2 = 0.559\ \ SE = 0.472\ \ F_{(1,\ 34)} = 9.733$$

$$- \log ki = 0.731\ (\pm 0.412)\ MR + 0.237 \qquad \rightarrow (2)$$

$$n = 36\ \ r^2 = 0.596\ \ SE = 0.412\ \ F_{(1,\ 34)} = 11.641$$

$$- \log ki = 0.412\ (\pm 0.246)\ Vw + 0.210 \qquad \rightarrow (3)$$

$$n = 36\ \ r^2 = 0.628\ \ SE = 0.430\ \ F_{(1,\ 34)} = 16.322$$

$$- \log ki = 0.578\ (\pm 0.836)\ Xeq - 0.366 \qquad \rightarrow (4)$$

$$n = 36\ \ r^2 = 0.196\ \ SE = 0.661\ \ F_{(1, 34)} = 0.310$$

M Log P, MR and Vw gave fairly good monoparametric regression equations. Equations (1), (2) and (3) corresponding to these parameters accounts for 56%, 60% and 63% variance ratio (R = 0.75, R = 0.77 & R = 0.79) F-values are also significant at 95% confidence level. Correlation of pki with Xeq, however, gave very poor regression equation. Equation (4) is statistically insignificant, having very low r^2 - value, high SE-value and low F-value. The coefficient of Xeq has very large value of error.

It may be noticed from **Table 3.2.1.B** that no autocorrelation exists between M Log P, MR and Vw so these parameters can be taken together. Performing MRA with M log P and MR gave the following equations.

$$- \log ki = 0.560\ (\pm 0.312)\ M\ Log\ P + 0.513\ (\pm 0.296)\ MR + 0.440 \qquad \rightarrow (5)$$

$$n = 36\ \ r^2 = 0.673\ \ SE = 0.412\ \ F_{(2, \ )} = 12.312$$

Equation (5) accounts for 67% (R = 0.82) variance ratio and F-value is significant at 95% level. Incorporation of Vw descriptor in equation (5) led to significant improvement in r^2 value.

$$- \log ki = 0.612\ (\pm 0.313)\ M\ Log\ P + 0.482\ (\pm 0.311)\ MR + 0.416\ (\pm 0.200)\ Vw + 0.332 \qquad \rightarrow (6)$$

$$n = 36\ \ r^2 = 0.757\ \ SE = 0.410\ \ F_{(3, \ )} = 18.276$$

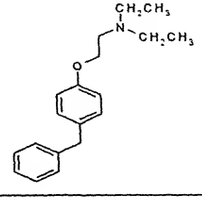
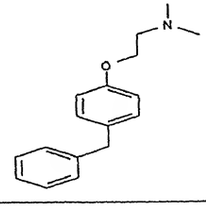
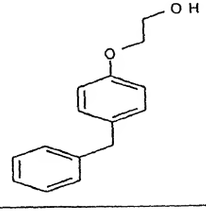
In our attempt to increase the correlation to a still higher value an indicator parameter Ind was used for the substituents having -N (C_4H_8) O group at the R position. A significant improvement in the r^2 value was observed.

$$- \log ki = 0.733\ (\pm 0.322)\ M\ Log\ P + 0.551\ (\pm 0.243)\ MR + 0.482\ (\pm 0.220)\ Vw - 0.182\ (\pm 0.220)\ Ind + 0.317 \rightarrow (7)$$

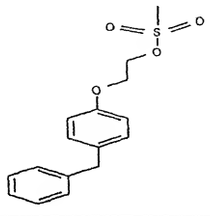
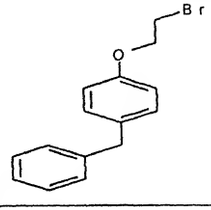
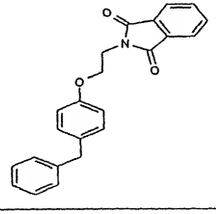
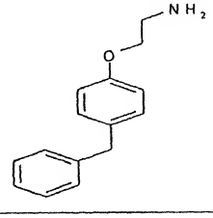
$n = 36\ r^2 = 0.878\ SE = 0.326\ F_{(4,\ )} = 28.340$

Equation (7) is sound in statistical parlance. It accounts for 88% variance ratio and the F-value is significant at 99% confidence level.

TABLE-3.2.1.A : PHYSICO-CHEMICAL DATA FOR NEW LIGANDS FOR THE MICROSOMAL ANTIESTROGEN BINDING SITE

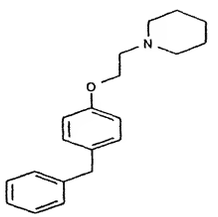
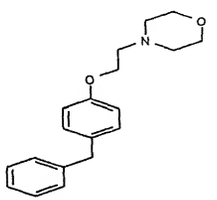
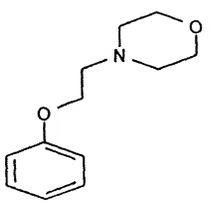
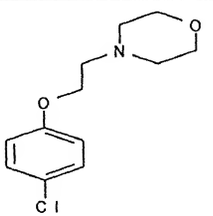
S.NO.	COMPD	STRUCTURE	DESCRIPTORS				OBSERVED ACTIVITY		CALCULATED ACTIVITY (Eq.7)	RESIDUAL
			MLOGP	MR	Xeq	Vw	Ki	LOG Ki		
1.	1		3.73	76.69	20.21	21.49	55.6	-1.75	-1.65	-0.10
2.	2		3.34	67.79	20.21	21.49	42.3	-1.63	-1.65	0.02
3.	3		3.18	70.52	20.21	21.49			-1.65	

[92]

4.	4		3.57	70.41	21.21	22.49			-1.65	
5.	5		2.50	70.10	20.72	22.82			-2.11	
6.	6		1.01	44.18	13.72	15.82			-3.69	
7.	7		1.28	49.45	14.75	17.09	924	-2.97	-2.90	-0.06

8.	8		1.41	52.27	15.10	17.00	1530	-3.18	-2.90	-0.28
9.	9		1.55	57.05	15.45	16.84	114	-2.06	-2.90	0.85
10.	10		1.28	46.81	14.72	16.82	8.8	-0.94	-1.46	0.51
11.	11		2.05	54.67	17.72	19.82			-1.32	

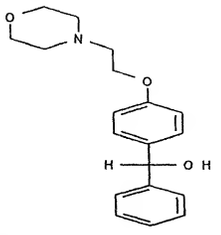
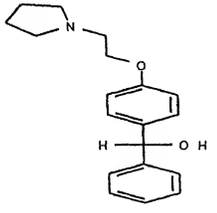
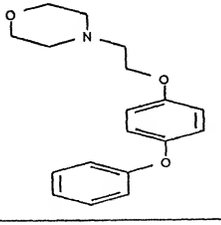
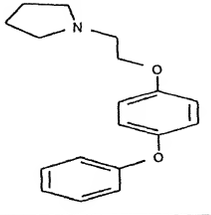
[94]

12.	12		3.26	64.13	18.21	19.49	28.1	-1.45	-1 65	0.20
13.	13		2.98	75.34	22.72	24.82	17.6	-1.25	-0.53	-0.72
14.	14		3.79	73.04	22.21	23.49			-0.07	
15.	15		1.54	68.92	21.23	24.16			-3.02	

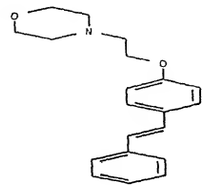
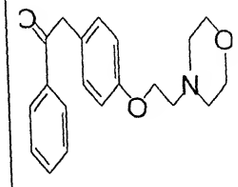
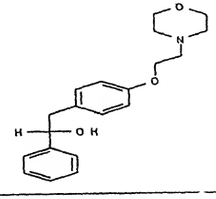
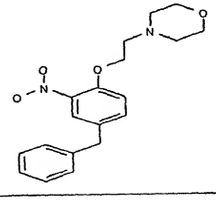
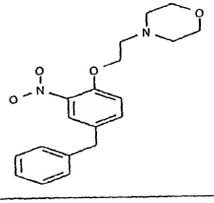
16.	16		2.35	66.62	20.72	22.82	3820	-3.58	-2.55	-0 30
17.	17		1.70	70.88	21.53	25.10	1630	-3.21	-2.91	-0.74
18.	18		2.50	68.57	21.02	23.77			-2.01	
19.	19		1.46	68.85	20.23	23.16	248	-2.39	-2.57	0.18

20.	20a		2.27	66.55	19.72	21.82	48.1	-1.68	-2.11	0 43
21.	20b		2.65	75.38	21.72	23.82	10.1	-1.00	-0.67	-0.34
22.	21a		1.84	71.54	22.23	25.16	102	-2.01	-2.18	0.17
23.	21b		1.92	73.50	22.53	26.10	64.5	-1.81	-1.03	-0.78

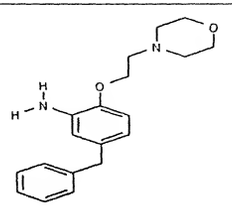
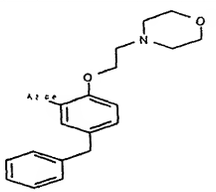
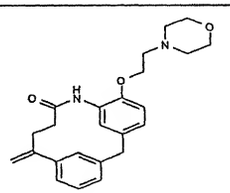
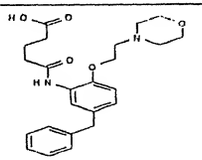
[97]

24.	22a		2.15	72.61	22.44	26.65			-2.01	
25.	22b		1.60	74.97	22.80	26.31			-1.11	
26.	23a		1.82	72.61	17.66	15.62	2.1	-0.32	-1.11	0.79
27.	23b			32.29	17.66	15.91	0.8	0.10	-0.51	0.61

[86]

28.	24		1.61	80.04	23.93	27.32	2010	-3.30	-3.22	-0.08
29.	25		1.25	87.60	26.95	31.98	850	-2.93	-3.44	0.51
30.	26		1.46	90.22	27.95	32.98			-3.44	
31.	27		4.02	65.36	18.90	17.50				
32.	28		3.11	63.22	22.33	16.00	80	-1.90	-2.28	0.37

[66]

33.	29		2.98	91.18	25.23	28.16	3	-0 48	-1.54	1 06
34.	30		2.77	56.63	16.21	17.49	730	-2.86	-2.28	-0.59
35.	31		2.75	63.77	18.90	20.66			-1.49	
36.	32		-0.18	63.54	19.21	20.49			-0.70	

CORRELATION MATRIX

TABLE: 3.2.1.B

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.230	1.000		
Vw	0.414	0.382	1.000	
Xeq	0.717	0.616	0.542	1.000

3.2.2 QSAR STUDIES ON 2-AMINO-4,6-DIARYLPYRIDINES AS NOVEL LIGANDS FOR THE ESTROGEN RECEPTOR[176].

Compounds in this series include 2-Amino-4, 6-diarylpyridines as novel ligands for the estrogen receptor effective in the treatment for both menopausal symptoms and for the prevention and management of postmenopausal osteoporosis. These ligands are developed as structurally novel templates which are readily amenable to parallel synthesis. Henke et. al[176] have evaluated the binding affinity of ERα and ERβ subtypes via a scintillation proximity assay (SPA) using a bacterial lysate containing over expressed GST-h ERα or GST-hERβ ligands binding domain.

The binding affinity data expressed as - log ki was correlated with hydrophobic (M log P), steric (MR and Vw) and electronic (Xeq) descriptors. **Table 3.2.2.A** lists the pki values along with the physicochemical data for a set of 16 substituted diarylpyridines for both ERα and ERβ subtypes for regression analysis. The correlation between the parameters for ERα and ERβ sybtypes used in the present study is recorded in **Table 3.2.2.B** and **Table 3.2.2.C**.

3.2.2.AQSAR STUDY OF SUBSTITUTED DIARYLPYRIDINES BINDING TO ERα SUBTYPES.

Simple linear regression analysis of M Log P, MR, Vw and Xeq with -log ki ERα gave the following correlations:

$$- \log ki = 0.943\ (\pm 0.428)\ M\ Log\ P - 0.853 \qquad \rightarrow (1)$$
$$n = 16\ \ r^2 = 0.571\ \ SE = 0.303\ \ F_{(1,\ 14)} = 4.440$$
$$- \log ki = 0.832\ (\pm 0.414)\ MR + 0.541 \qquad \rightarrow (2)$$
$$n = 16\ \ r^2 = 0.326\ \ SE = 0.573\ \ F_{(1,\ 14)} = 1.680$$
$$- \log ki = 0.636\ (\pm 0.312)\ Vw + 0.394 \qquad \rightarrow (3)$$
$$n = 16\ \ r^2 = 0.042\ \ SE = 6.930\ \ F_{(1,\ 14)} = 0.731$$

$$- \log ki = 0.414\ (\pm 0.208)\ Xeq + 0.611 \qquad \rightarrow (4)$$

$$n = 16 \quad r^2 = 0.003 \quad SE = 8.720 \quad F_{(1, 14)} = 0.050$$

The monoparametric equations (3) and (4) show very poor correlations. Equations (1) and (2) however can be accepted from statistical point of view. Their F-values are significant at 90% confidence interval. It may be noticed from **Table 3.2.2.B** that no autocorrelation exists between M log P and MR so these descriptors can be used to perform multiple regression analysis.

$$- \log ki = 0.860\ (\pm 0.424)\ M\ Log\ P + 0.697\ (\pm 0.301)\ MR - (0.819) \qquad \rightarrow (5)$$

$$n = 16 \quad r^2 = 0.709 \quad SE = 0.316 \quad F_{(2, 13)} = 11.680$$

Equation (5) is a fairly good regression equation. It accounts for 71% variance ratio (R = 0.84) and the F-value is also significant at 99% confidence level.

In our ongoing efforts to improve the correlation to a still higher level an indicator parameter, Ind, for the number of acceptor atoms for H - bonds (N, O, F) was taken. It was considered one when the number of acceptor atoms for H-bonds was three and zero in the remaining cases. Significant improvement in correlation was observed on incorporating, Ind, in equation (5) which is shown below:

$$- \log ki = 0.863\ (\pm 0.332)\ M\ Log\ P + 0.419\ (\pm 0.198)\ MR + 0.102\ (\pm 0.007)\ Ind + 0.199 \qquad \rightarrow (6)$$

$$n = 16 \quad r^2 = 0.922 \quad SE = 0.213 \quad F_{(3, 12)} = 24.780$$

Equation (6) is excellent in statistical parlance. It accounts for 92% variance ratio (R = 0.96) and F-value is significant at 99% confidence interval.

3.2.2.BQSAR STUDY OF SUBSTITUTED DIARYLPYRIDINES BINDING TO ERβ SUBTYPE.

The correlation of the binding affinity for ERβ subtype with hydrophobic, steric and electronic descriptors gave the following regression equations:

$$-\log ki = 0.516\ (\pm 0.268)\ M\ Log\ P - 0.362 \quad \rightarrow (7)$$
$$n = 16 \quad r^2 = 0.374 \quad SE = 0.512 \quad F_{(1,\ 14)} = 1.769$$
$$-\log ki = 0.648\ (\pm 0.581)\ MR + 0.346 \quad \rightarrow (8)$$
$$n = 16 \quad r^2 = 0.108 \quad SE = 0.602 \quad F_{(1,\ 14)} = 0.321$$
$$-\log ki = 0.521\ (\pm 0.411)\ Vw + 0.414 \quad \rightarrow (9)$$
$$n = 16 \quad r^2 = 0.116 \quad SE = 0.577 \quad F_{(1,\ 14)} = 0.365$$
$$-\log ki = 0.662\ (\pm 0.683)\ Xeq + 0.312 \quad \rightarrow (10)$$
$$n = 16 \quad r^2 = 0.008 \quad SE = 7.261 \quad F_{(1,\ 14)} = 0.041$$

The above monoparametric equations shows very poor correlations with the binding affinity values. From the correlation matrix (**Table 3.2.2.C**) it was observed that all the descriptors showed high value of autocorrelations amongst themselves except M Log P and Vw wherein no autocorrelation was seen. Hence MRA was performed taking these two descriptors together and the following correlation was obtained

$$-\log ki = 0.761\ (\pm 0.502)\ M\ Log\ P + 0.613\ (\pm 0.343)\ Vw + 0.187 \rightarrow (11)$$
$$n = 16 \quad r^2 = 0.774 \quad SE = 0.512 \quad F_{(2,\ 13)} = 8.601$$

Equation (11) accounts for 77% variance ratio and F-value is significant at 95% confidence level. Although equation (11) is satisfying from statistical point of view their exists a lacunae in it that the error terms in this equation is comparatively higher. In order to remove this lacunae and improve the degree of correlation an indicator parameter, Ind-1 for the number of rotable bonds was taken. It was assigned the value-1 when the number of rotable bonds was equal to or greater than five (5) and zero in other cases. A significant improvement in correlation was observed:

$$- \log ki = 0.431\ (\pm 0.120)\ M\ Log\ P + 0.368\ (\pm 0.107)\ Vw + 0.132\ (\pm 0.004)\ Ind1 + 0.214 \quad \rightarrow (12)$$

$$n = 16 \quad r^2 = 0.848 \quad SE = 0.322 \quad F_{(3,\ 12)} = 16.330$$

Equation (12) accounts for 85% variance ratio (R = 0.92) and F-value is significant at 99% confidence level. This equation is very well acceptable from statistical point of view.

TABLE-3.2.2.A : PHYSICO-CHEMICAL DATA FOR 2-AMINO-4, 6-DIARYLPYRIDINES

S. No.	COMPD	SUBSTITUENTS				DESCRIPTORS				OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResERαKi	OBSERVED ACTIVITY		CALCULATED ACTIVITY	ResERβKi
		R^1	R^2	R^3	R^4	MLogP	MR	Xeq	Vw	ERαKi	LogERαKi	PreERαKi Eq. (6)		ERβKi	LogERβKi	PreERβKi Eq. (12)	
1.	1	H	H	$-CH_3$	$-(CH_2)_2$-Ph	3.06	95.01	30.82	28.6	490	-2.69	-2.69	0	1050	-3.021	-3.09	0.07
2.	3	H	o-CH_3	$-CH_3$	$-(CH_2)_2$-Ph	3.09	95.87	30.82	28.6	650	-2.813	-2.69	-0.12	1100	-3.041	-3.09	0.05
3.	4	m-CH_3	H	$-CH_3$	$-(CH_2)_2$-Ph	3.26	98.29	31.82	29.6	160	-2.204	-2.69	0.49	680	-2.833	-2.98	0.15
4.	5	H	H	$-CH_3$	$-(CH_2)$-CH-$(CH_3)_2$	3.29	98.29	31.82	29.6	620	-2.792	-2.69	-0.10	1260	-3.1	-3.09	-0.01
5.	6	m-CH_3	H	$-CH_3$	$-(CH_2)$-CH-$(CH_3)_2$	3.26	98.85	31.82	29.6	210	-2.322	-2.69	0.37	660	-2.82	-2.98	0.16
6.	7	m-CH_3	H	$-CH_3$	$-(CH_2)_2-CH_3$	5.01	125.04	36.66	34.9	660	-2.82	-2.69	-0.13	790	-2.898	-2.98	0.09
7.	9	m-CH_3	H	$-CH_3$	$-CH_2$-Ph	2.48	98.17	32.32	28.93	380	-2.58	-2.69	0.11	1120	-3.049	-2.98	-0.07
8.	10	m-CH_3	H	$-CH_3$	$-CH_2$-(1-naphthyl)	3.9	98.47	29.66	27.9	450	-2.653	-2 69	0.04	1410	-3.149	-3.53	0.38
9.	11	m-CH_3	H	$-CH_3$	$-(CH_2)_2$-N-$(CH_3)_2$	4.03	101.09	30.66	28.9			-2.69				-2 98	
10.	12	m-CH_3	H	$-CH_3$	$-(CH_2)_2$-(o-pyridyl)	4.06	101.09	30.66	28.9	1260	-3.1	-3.064	-0.04			-2.98	
11.	13	H	H	-$(CH_2)_2$ CH_3	$-CH_2$-(o-pyridyl)	3.27	80.42	25.66	23.9	180	-2.255	-2.05	-0.21	710	-2 851	-2.74	-0.11

[106]

12.	14	m-CH_3	H	-$(CH_2)_2$ CH_3	-CH_2-(o-pyridyl)	3.45	83.04	26.66	24 9	70	-1.845	-2.05	0.21	330	-2 519	-2 63	0.11
13.	15	o-CH_3	H	-$(CH_2)_2$ CH_3	-CH_2-(o-pyridyl)	3.66	85.66	27.66	25.9			-2.05				-2 63	
14.	16	m-CH_3	H	- CH_2C H_3	-CH_2-(p-pyridyl)	3.86	98.47	29.66	27.9	830	-2.919	-2.956	0.04	1000	-3	-2.9	-0.1
15.	17	m-CH_3	H	- CH_2P h	-$(CH_2)_2$-Ph	4.91	113.99	33.66	31.9	130	-2.114	-2.126	0 01	250	-2.398	-2 53	0.13
16.	18	m-CH_3	H	-$(CH_2)_2$ OCH_2 CH_3	-$(CH_2)_2OCH_2CH_3$	2.64	85.65	27.82	25.6	2000	-3.301	-2.33	-0.97	1170	-3.068	-3 01	-0.05

CORRELATION MATRIX

TABLE: 3.2.2.B

(FOR ERα SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.112	1.000		
Vw	0.431	0.212	1.000	
Xeq	0.663	0.814	0.720	1.000

CORRELATION MATRIX

TABLE: 3.2.2.C

(FOR ERβ SUBTYPE)

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.562	1.000		
Vw	0.386	0.986	1.000	
Xeq	0.762	0.721	0.562	1.000

3.2.3 QSAR STUDIES ON CIS-3, 4-DIARYL-HYDROXY CHROMANES AS HIGH AFFINITY PARTIAL AGONISTS FOR ESTROGEN RECEPTOR[177].

Compounds in this series include, the cis-3, 4-diaryl-hydroxy-chromanes, a new group of non-steroidal partial estrogens effective in the treatment of postmenopausal degenerative diseases particularly osteoporosis and coronary heart disease. The IC_{50} binding affinities of the chromanes to the estrogen receptor were determined by Bury et al[117] by measuring their ability to compete with [^{3}H]-17β-estradiol for receptor binding in ER-rich cytosol derived from rabbit uterine tissue in dextran coated charcoal (DCC) assay [178, 179]. In this discussion QSAR have been performed on the racemate and enantiomeric forms of substituted hydroxy chromanes using hydrophobic (M Log P), steric (MR and Vw) and electronic (Xeq) descriptors.

Table 3.2.3.A lists the log IC_{50} values along with the physicochemical data for a set of 16 substituted racemate and enantiomeric hydroxy chromanes. **Table 3.2.3.B** and **Table 3.2.3.C** represents the correlation matrices between the descriptors used in the present study for racemate and enantiomeric hydroxy chromanes.

3.2.3.AQSAR STUDY OF RACEMATE FORM OF SUBSTITUTED HYDROXY CHROMANES

Simple linear regression analysis of M Log P, MR, Vw and Xeq with log IC_{50} gave the following correlations:

$$- \log IC_{50} = 0.926\ (\pm 0.731)\ M\ Log\ P + 0.396 \quad \rightarrow (1)$$
$$n = 16\ \ r^2 = 0.282\ \ SE = 0.552\ \ F_{(1,\ 14)} = 0.461$$
$$- \log IC_{50} = 0.822\ (\pm 0.631)\ MR + 0.712 \quad \rightarrow (2)$$
$$n = 16\ \ r^2 = 0.116\ \ SE = 0.612\ \ F_{(1,\ 14)} = 0.417$$
$$- \log IC_{50} = 0.722\ (\pm 0524)\ Xeq - 0.315 \quad \rightarrow (3)$$
$$n = 16\ \ r^2 = 0.327\ \ SE = 0.414\ \ F_{(1,\ 14)} = 0.826$$
$$- \log IC_{50} = 0.811\ (\pm 0.674)\ Vw + 0.540 \quad \rightarrow (4)$$
$$n = 16\ \ r^2 = 0.045\ \ SE = 0.686\ \ F_{(1,\ 14)} = 0.181$$

Equation (1), (2), (3) and (4) show very poor correlation with the binding affinity. After checking the correlation matrix **(Table 3.2.3.B)** it was observed that no autocorrelation exists between M Log P and MR and also M Log P and Xeq. so these parameters were taken together and multiple regression analysis was performed which yielded the following regression equations:

$$-\log IC_{50} = 0.932\ (\pm 0.814)\ M\ Log\ P + 0.674\ (\pm 0.696)\ MR + 0.611 \rightarrow (5)$$

$$n = 16\ \ r^2 = 0.374\ \ SE = 0.482\ \ F_{(2, 13)} = 0.621$$

$$-\log IC_{50} = 1.011\ (\pm 0.622)\ M\ Log\ P - 0.321\ (\pm 0.131)\ Xeq + 0.410 \rightarrow (6)$$

$$n = 16\ \ r^2 = 0.634\ \ SE = 0.426\ \ F_{(2, 13)} = 5.440$$

Significant improvement in correlation was observed when an indicator parameter, Ind was introduced in equation (6). Indicator parameter has been taken for meta substituents at G position the value for which is taken as unity if the substituent is attached at the meta position of the aromatic ring and zero in all the other cases.

$$-\log IC_{50} = 0.523\ (\pm 0.571)\ M\ Log\ P - 0.511(\pm 0.312)\ Xeq + 0.230 + 0.311\ (\pm 0.102)\ Ind + 0.414 \rightarrow (7)$$

$$n = 16\ \ r^2 = 0.870\ \ SE = 0.246\ \ F_{(3, 12)} = 20.450$$

Equation (7) accounts for 87% variance ratio (R = 0.93) and F-value is also significant at 99% confidence interval.

3.2.3.BQSAR STUDY OF ENANTIOMERIC FORM OF SUBSTITUTED HYDROXY CHROMANES

Simple regression analysis using M log P, MR, Xeq and Vw with - log IC_{50} resulted in the following regression equations :

$$-\log IC_{50} = 0.812\ (\pm 0.664)\ M\ log\ P + 0.414 \rightarrow (8)$$

$$n = 16\ \ r^2 = 0.252\ \ SE = 0.447\ \ F_{(1, 14)} = 0.474$$

$$-\log IC_{50} = 0.761\ (\pm 0.510)\ MR + 0.556 \rightarrow (9)$$

$$n = 16\ \ r^2 = 0.198\ \ SE = 0.612\ \ F_{(1, 14)} = 0.248$$

$-\log IC_{50} = 0.861\ (\pm 0.662)\ Xeq - 0.532 \rightarrow (10)$

$n = 16\ r^2 = 0.246\ SE = 0.510\ F_{(1, 14)} = 0.326$

$-\log IC_{50} = 0.961\ (\pm 0.772)\ Vw - 0.661 \rightarrow (11)$

$n = 16\ r^2 = 0.046\ SE = 0.718\ F_{(1, 14)} = 0.016$

From equation (8), (9), (10) and (11) it is observed that the linear regression gave very poor correlations. After checking the correlation matrix **(Table 3.2.3.C)** it can be inferred that no autocorrelation exists between M Log P and MR and M Log P and Xeq. So these parameters can be taken together. On applying multiple regression analysis, following regression equation has been obtained :

$-\log IC_{50} = 0.574\ (\pm 0.221)$ M Log P $+ 0.592\ (\pm 0.331)$ MR $+ 0.772 \rightarrow (12)$

$n = 16\ r^2 = 0.515\ SE = 0.441\ F_{(2, 13)} = 2.316$

$-\log IC_{50} = 0.769\ (\pm 0.513)$ M Log P $- 0.661\ (\pm 0.512)$ Xeq $+ 0.512 \rightarrow (13)$

$n = 16\ r^2 = 0.447\ SE = 0.417\ F_{(2, 13)} = 1.761$

Moderate improvement in r^2 values is observed in equations (12) and (13). These equation however are not very sound in statistical parlance as they have high standard error values and F-values are significant at 90% confidence level. In an attempt to obtain a statistically sound regression equation an indicator parameter, Ind 1, denoting the presence of fluorine atom was introduced. As a consequence excellent improvement in correlations was observed which is depicted by the following equations :

$-\log IC_{50} = 0.612\ (\pm 0.226)$ M Log P $+ 0.586\ (\pm 0.210)$ MR $+ 0.018\ (\pm 0.007)$ Ind 1 $+ 0.310 \rightarrow (14)$

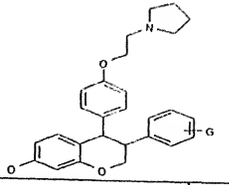
$n = 16\ r^2 = 0.836\ SE = 0.263\ F_{(3, 12)} = 18.330$

$-\log IC_{50} = 0.712\ (\pm 0.288)$ M Log P $- 0.513\ (\pm 0.310)$ Xeq $+ 0.197\ (\pm 0.009)$ Ind 1 $+ 0.312 \rightarrow (15)$

$n = 16\ r^2 = 0.763\ SE = 0.301\ F_{(3, 12)} = 11.760$

Equations (14) and (15) account for 84% and 76% variance ratio ($R = 0.91$ and $R = 0.87$) and F-value is also significant at 99% confidence interval.

TABLE-3.2.3.A : PHYSICO-CHEMICAL DATA FOR HYDROXY-CHROMANES



S. No.	COMPD	G	SIDE CHAIN		DESCRIPTORS				ER BINDING		OBSERVED (Eq.7)	RESIDUAL	ER BINDING		OBSERVED (Eq.15)	RESIDUAL
			n	m	M LogP	MR	Xeq	Vw	IC_{50} (RACEMATE)	-Log IC_{50}			IC_{50} (ENANTIOMER)	-Log IC_{50}		
1	a	H	1	1	3.188	104.809	35.99	31.13	18	-1.26	-0.91	-0.35	400	-2.6	-2.95	0.35
2	b	para-F	1	1	3.188	101.928	33.16	30.23	5	-0.7	-0.91	0.21	1500	-3.18	-2.95	-0.23
3	c	para-CF_3	1	1	3.384	104.543	34.16	31.23	31	-1.49	-1.91	0.42	410	-2.61	-2.89	0.27
4	d	para-OH	1	1	3.672	108.362	38.53	32.46	2	-0.3	-0.42	0.12	140	-2.15	-2.06	-0.03
5	e	para-Me	1	1	3.348	107.430	36.99	32.13	12	-1.08	-1.01	-0.07	1100	-3.04	-2.89	-0.16
6	f	para-Phenyl	1	1	3.480	105.225	35.61	31.64	55	-1.74	-1.42	-0.32			-2.84	
7	g	meta-CF_3	1	1	2.596	104.425	35.49	31.74	13	-1.11	-0.7	-0.42	550	-2.74	-2.78	0.04
8	h	meta-F	1	1	3.089	99.984	33.61	29.64	1	0	0.53	-0.63	1700	-3.23	-2.94	-0.29
9	i	meta-OH	1	1	3.089	99.984	33.61	29.64	0	2	0.53	1.47	160	-2.2	-2.29	0.08
10	j	meta-Me	1	1	3.480	105.742	37.53	31.46	0	0.7	0.52	0.17	160	-2.2	-2.29	0.08
11	k	H	1	2	2.202	99.184	33.49	29.74	3	-0.48	-0.44	-0.04	85	-1.93	-1.93	0
12	l	H	2	2	3.188	104.809	35.99	31.13	2	-0.3	-0.91	0.64	100	-2	-2	0
13	m	para-CF_3	1	2	3.931	122.60	38.16	35.23	23	-1.36	-1.4	0.04	360	-2.56	-2.63	0.07
14	n	para-Me	1	2	3.480	105.724	37.53	31.46	3	-0.48	-0.32	-0.16	1300	-3.11	-2.76	-0.35
15	o	para-F	2	2	3.089	99.984	33.61	29.84	6	-0.78	-0.32	-0.46	800	-2.9	-2.94	0.04
16	p	meta-OH	2	2	2.202	99.184	33.49	29.74	2	-0.3	0.4	-0.7			-2 23	

CORRELATION MATRIX

TABLE: 3.2.3.B

RACEMATE FORM

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.482	1.000		
Vw	0.747	0.661	1.000	
Xeq	0.226	0.559	0.881	1.000

TABLE: 2.2.3.B

ENANTIOMERIC FORM

	M Log P	MR	Vw	Xeq
M Log P	1.000			
MR	0.316	1.000		
Vw	0.592	0.404	1.000	
Xeq	0.722	0.869	0.881	1.000

CONCLUSIONS

From the regression analysis data of the various series (3.2.1, 3.2.2, 3.2.3) studied in this section for the binding affinity of estrogen receptors expressed in IC_{50} (nM) forms, it is observed that hydrophobic descriptor M Log P here also has a significant role to play however in this case M Log P enhances the binding as its coefficient value is positive reverse from the previous section. Besides MR and Xeq, Vw here has positive contribution towards the binding affinity. Investigation of the various non-steroidal analogues taken for QSAR study in this section reveals that sterically large, non-polar groups are necessary to achieve good receptors binding affinity.

From the correlations of AEBS binding affinities of diphenyl methane based tamoxifen derivatives (Series 3.2.1) that are selective for AEBS over the estrogen receptor M Log P, MR and Vw descriptors are critical for the affinity to AEBS. Good linear quantitative structure affinity relationship has been established. Equation (6) and (7) (Series 3.2.1) illustrates that the rigid triphenylethylenic moiety of tamoxifen defines a spatial relationship driving the occupancy of the tamoxifen within AEBS and the binding of compounds. The positive coefficient of indicator parameter; indicates that morpholinic derivatives have higher affinity. This illustrates that the ring part of the amine is important for high affinity.

In series 3.2.2 and 3.2.3 positive hydrophobic term appears in both ERα and ERβ subtypes, indicating that their hydrophobic contribution to the binding with the substituents examined for diarylpyridines as well as hydroxy chromanes. Most of the substituents however are relatively polar and would not provide much opportunity for enhanced hydrophobic bonding. The positive coefficient of Vw is also significant because it will result in raising the binding affinity values for bulkier molecules. The positive coefficient of indicator parameter suggests that the substituents having the tendency to accept H-bonds are preferred in enhancing the binding affinity.

The common feature observed in all the series was that M Log P has large and positive coefficient which indicates that hydrophobic substituents are favoured. The positive coefficient of MR indicates, that sterically bulky substituents raise the binding affinity with these parameters, when the substituents become larger, the positive steric interaction becomes dominant and affinity increases. This view is consistent with the crystal structure of the estrogen receptor ligand binding domain[180].

The X-ray crystal structure of the receptor shows that there is some very hydrophobic space above the B-ring of the ligand at position 11, presumably sufficient to accommodate groups of moderate size without contacting the binding surface of the receptor. Larger substituents would require some movement of ligand or receptor for a complex to form. The conclusions from QSAR of the non-steroidal compounds performed in this section are similar to those made from an earlier analysis of structure affinity relationship.

Thus as a generalization substituents that increase the electron density on the phenolic ring appear to increase the binding affinity. Also from the positive hydrophobic interaction between ligand substituents and receptor (the same phenomenon has been shown in the CoMFA analysis[181] needs ot be tempered by the recognition that the ABCD tetracyclic core structure of steroidal estrogens (as well as the corresponding units in nonsteroidal estrogens) is generally very hydrophobic and may contribute to the bulk of ligand binding by a hydrophobic mechanism.

CHAPTER-4

3D-QSAR AND MOLECULAR MODELLING

SECTION-A

INTRODUCTION TO 3D QSAR USING APEX-3D AND CATALYST

(I) *APEX-3D* : AN EXPERT SYSTEM FOR AUTOMATED PHARMACOPHORE IDENTIFICATION *AND* 3D-QSAR MODEL BUILDING

The calculation of 3D-QSAR models was done on which is developed to represent, elucidate and utilize knowledge on structure activity relationships. Apex-3D can be used to build 3D-SAR and 3D-QSAR models which can be used for activity classification and prediction. The general principle of operation is based on emulating the intelligence of the researcher engaged in establishing relationships between a compound's structural parameters and its activity. The corner stone of the Apex-3D methodology is automated identification of biophores (pharmacophores). Theses biophores can be used for building qualitative activity prediction rules and for creating search queries to identify new leads in a 3D-database. Identified biophores can be used as starting points for constructing 3D-QSAR models when good quantitative data is available. Combination of a 3D pharmacophore with a quatitative regression equation is unique to the Apex-3D approach. Prediction of activity for novel compounds requires the biophore be present. The activity level is calculated from QSAR equation.

4.1 GENERAL PRINCIPLES OF APEX-3D: IDENTIFICATION OF BIOPHORIC PATTERNS AND BUILDING 3D-QSAR MODELS

Apex-3D is based on the logico-structural approach to drug design developed by Dr. Valery Golender and his colleagues (1980, 1983, 1993, 1995). This approach, to a certain extent, simulates the intelligence of a scientist engaged in establishing relationships between certain structural characteristics of compounds and their activity.

These basic inductive methods of agreement, difference, and concomitant variations are used by researchers to identify structural patterns associated with bioactivity:

(1) The agreement method is based on identification of the common structural patterns in different compounds having the same type of biological activity.

(2) The difference method is based on identification of the different structural patterns in active and inactive compounds.

(3) The concomitant variations method is based on identification of variation in structural properties that explain changes in the biological activity of a set of compounds.

4.1.1 PRINCIPLE OF AUTOMATED IDENTIFICATION OF BIOPHORIC PATTERNS

The automated identification of biophoric patterns according to the logico-structural approach is based on the agreement and difference methods and involves the following steps:

1. Separation of data set compounds into activity classes according to their activity type or level.

2. Generation of structural representations based on topological (2D) or topographical (3D) distance matrices, and sets of structural indexes for identifying biophoric descriptor centers in chemical compounds.

3. Identification of common structural patterns (features) in all pairs of compounds belonging to a common activity class.

4. Calculation of the number of occurrences of all identified structural patterns (features) among compounds from each activity class of the analyzed data sets.

These occurrence numbers are used to calculate statistical estimates of features:

1. The probability that novel compounds having a given feature will belong to a certain activity class.

2. The reliability calculated as the probability of nonchance occurrence of the feature.

3. Identification of biophores. Biophores are selected as features having both probability and reliability higher than certain thresholds. These thresholds are established during training of the activity prediction system.

4. Prediction of biological activity of novel compounds which have been synthesized, or suggested for synthesis, based on the identified biophores.

5. Analysis of computer selected biophores and application to the rational synthesis of compounds possessing desirable biological activity.

4.1.2 PRINCIPLE OF 3D-QSAR MODEL GENERATION

The identified biophores are used as starting points for building 3D-QSAR models. This procedure parallels reasoning based on the concomitant variations method and involves the following steps:

1. Automated identification of biophores (1 through 5 above).
2. Optimization of the superimposition of all compounds containing a selected biophore.
3. Building 3D-QSAR models for the selected biophores based on correlation of ligand active sites and global molecular properties with biological activity.

4. Prediction of activity level of novel compounds based on identified biophores and 3D-QSAR models.

5. Analysis of selected 3D-QSAR models and application to rational drug design.

4.2 APEX-3D STRUCTURAL DESIGN: MODULES AND CONCEPTS

Apex-3D is a learning rule-based expert system that accumulates knowledge and makes inferences on structure-activity relationships from structure-activity data. The system stores two types of information:

1. Data on topological (2D) and topographical (3D) structures of chemical compounds.

2. Knowledge of structure-activity relationships expressed as rules that associate biophoric patterns with particular biological activities.

Apex-3D rules take the following forms:

Qualitative rules:

IF structure S contains the biophoric pattern B_i, THEN it possesses the activity A_k with probability Pik.

Quantitative rules:

IF structure S contains the biophoric pattern B_i, having the associated QSAR model $A_k = F(B_j, S)$, THEN it possesses the activity A_k calculated according to the model.

The main component of Apex-3D is an inference engine containing two modules:

An inductive inference engine used for biophore search and rule generation.

A deductive inference engine used for rule-based activity prediction.

Rule generation is based on structure matching procedures of the logico-structural approach.

4.2.1 APEX-3D MODULES

4.2.1.1 INSIGHT II COMPONENT

Apex-3D is integrated with the Insight II molecular modelling and molecular graphics environment in such a way that one can:

Build 2D sketches and 3D models of chemical compounds in Insight II and use Apex-3D for activity prediction and biophore searching.

Define the **Task Definition** parameters interactively using Insight II tables.

Export compounds, biophores, and superimpositions of compounds sharing a common biophoric pattern from Apex-3D to Insight II.

4.2.1.2 COMPUTATIONAL CHEMISTRY COMPONENT

Computational chemistry programs calculate structural indexes used for biophore definition and conformational space clustering. These programs include:

- MOPAC 6.0 (QCPE program) postprocessor for calculating MOPAC-based indexes, including atomic charges, pi-populations, and donor and acceptor indexes.
- Module for calculating hydrophobicity and molar refractivity based on atomic contributions (Ghose et al. 1988; Viswanadhan et al. 1989).
- Module for calculating indexes based on qualitative models. A detailed description of indexes is given in.
- Module for reducing the number of conformations using a clusterization methodology.

4.2.2 BIOPHORE CONCEPT OF APEX-3D

A biophore represents a certain structural and electronic pattern in a bioactive molecule which is responsible for its activity, possibly due to receptor interaction. Apex-3D identifies two types of biophores:

- Topological (2D) biophores based on graph-theoretical structure representation.
- Topographical (3D) biophores based on 3D structure representation.

4.2.2.1 TOPOLOGICAL BIOPHORES

Topological biophores consist of common descriptor centers separated by some number of bonds. These biophores are displayed for one molecule at a time. The bonds separating the descriptor centers can be highlighted. Note that multiple structures cannot be superimposed on the biophore.

4.2.2.2 TOPOGRAPHICAL BIOPHORES

Topographical biophores define certain superimpositions of compounds in 3D space according to their common biophores. This superimposition serves as a convenient graphical image, as well as a tool for extracting additional structural information on the chirality and environment of biophores. Since biophore identification in Apex-3D is based on distance matrices that are invariant under reflection, biophores are isomorphic for different stereoisomers. Thus the resulting superimposition of mirror images usually causes deviations in the positions of matched atoms.

4.3 MOLECULAR ALIGNMENT AND MOLECULAR SUPERIMPOSITION IN APEX-3D

Each identified biophore serves as a basis for molecular superimposition by defining a common coordinate system for molecules sharing that biophore. But for most compounds this superimposition is not unique because several molecular fragments can match the biophore and a number of conformers can fit it. Therefore, a special procedure for superimposition optimization is needed to select the best, or at least a reasonable, superimposition.

Molecular alignment in Apex-3D for superimposition optimization is based on three simple intuitive principles:

1. ***Biophore anchor principle:*** Biophoric centers must be superimposed for all molecules with minimal deviation.

2. ***Similarity principle:*** Atoms of the same chemical type from different molecules must be superimposed as closely as possible.

3. ***Atom adjacency principle:*** The closer an atom is to a biophoric center, the more important is its alignment.

4.4 CHEMICAL STRUCTURE REPRESENTATION IN APEX-3D: *DESCRIPTOR CENTERS AND DISTANCE MATRICES*

As mentioned earlier, a biophore represents a certain structural pattern presumed to be responsible for the biological activity. These biophores are the basis for Apex-3D's approach to chemical structure representation. There are two parts to biophore representation in Apex-3D:

1. ***Descriptor Centers*** that represent parts of hypothetical biophoric moieties capable of interacting with a receptor.

2. ***Distance Matrices*** that describe the mutual orientation of descriptor centers using topological (number of bonds) or topographical distances (angstroms).

Descriptor centers can be either atoms or pseudoatoms that can participate in the ligand-receptor interactions based on the following types of physical properties:

- Electrostatic interactions
- Hydrogen bonds
- Charge transfer complexes
- Hydrophobic interaction
- van der Waals (or London) dispersion forces

These physical properties correlate with certain structural indexes which are calculated using various computational chemistry methods

All of the descriptor center information is stored by Apex-3D in two matrices:

1. The ***Property Matrix*** stores the structural indexes for all descriptor centers identified in a given structure.
2. The ***Distance Matrix*** stores the distances between all pairs of descriptor centers.

Data from these matrices are used to define the biophores which are a subset of the matrices common to several compounds.

4.4.1 STRUCTURAL INDEXES FOR DESCRIPTOR CENTERS

In Apex-3D, structural indexes are divided into two groups:

1. Indexes calculated using the MOPAC 6.0 semi-empirical quantum-chemical program (QCPE program).

2. Indexes calculated using simplified computational chemistry models.

4.4.2 *DISTANCE MATRICES* ALGORITHM FOR BIOPHORE EXTRACTION IN APEX-3D: *EXHAUSTIVE SEARCH AND FAST SEARCH*

Structure matching in Apex-3D is based on the selection of maximal common 2D or 3D patterns of biophoric centers between two compounds. The algorithm constructs the *compatibility graph* and selects its cliques (Golender and Rosenblit 1983 pp. 129-143).

An efficient clique-finding procedure is used to select the compatibility graph cliques and delete isomorphic patterns.

Apex-3D supports two versions of the biophore extraction algorithm for compounds with multiple conformations.

1. *Exhaustive Search:* Matches all possible pairs of conformers of all compounds. Each conformer of each compound will be matched with each conformer of all the other compounds in the training set.

2. *Fast Search:* Matches the first conformer of a compound with all conformers of all the other compounds. This is done for each compound in the training set.

Both of these algorithms use the first compound as a template when setting biophore property and distance matrices. In situations where the names given to the training set of compounds places an unusual molecule as the first compound (template), Apex-3D may have trouble

identifying the best biophores for training. In these rare instances, a priority assignment can be made during task definition that will allow you to specify compound ordering

4.5 ACTIVITY PREDICTION TRAINING SYSTEM: PREDICTION OF BIOLOGICAL ACTIVITY ON THE BASIS OF IDENTIFIED BIOPHORES/ PHARMACOPHORES

The ultimate goal of biophore selection is to better design new bioactive compounds. One of the tools for achieving this goal is the prediction of biological activity on the basis of identified biophores.

Apex-3D's activity prediction system works as a filter, filtering out inactive compounds and sending only supposedly active compounds to the output.

Supposing there are compounds as input, of them being active. The activity prediction system recognizes n compounds as active, of them actually being active.

Two types of errors can be made during prediction. The first type is associated with missed active compounds. Active compounds classified as inactive are called ***false negatives***.

The second type of error is associated with classification of inactive compounds as active. These compounds are called ***false positives***.

Apex-3D uses two types of activity prediction training procedures:

1. ***Reclassification*** where prediction is done on one of the training set compounds using all of the training set compounds.

2. ***Leave-one-out recognition*** where training is done *N* times for *N* - 1 compounds of the data set, and a prediction is made for the *N*th compound.

Leave-one-out recognition gives a more realistic estimate of the predictive power of the recognition system. This is because one of the compounds is excluded from the training set and a prediction of activity is made based on a model generated from the remaining set of compounds. During the training process, Apex-3D automatically sets thresholds for the selection of biophores according to probability and reliability.

4.5.1 ACTIVITY PREDICTION

Apex-3D uses the rules generated during task building for activity prediction. New molecules for prediction can consist of one or more conformations just like the compounds used in the training set.

There are two modes of operation for activity prediction:

- **Classification**, in which Apex-3D attempts to assign a new molecule to one or more activity classes that were defined for the task.
- **Quantitative**, in which Apex-3D attempts to predict the activity of a new molecule based on 3D-QSAR models present in the task's knowledge base.

4.6 3D-QSAR MODELS IN TERMS OF BIOPHORIC SITES AND SECONDARY SITES

3D-QSAR model building in Apex-3D allows you to identify potential interactive sites in ligand molecules and correlate the physicochemical properties of these sites and global molecular properties with available quantitative biological data. Ligand active sites are centered on atoms and are divided into two groups:

- ***Biophore sites*** are centers of specific ligand-receptor interactions participating in biophore definition and present in all analyzed molecules.
- ***Secondary sites*** are centers of specific ligand-receptor interactions that may be present in only a subset of the analyzed structures and allow mapping of secondary receptor pockets which may modify ligand activity.

Such subdivision of active site groups can also tailor the complexity of the 3D-QSAR model towards available data. Models based only on biophore sites are more robust and less influenced by conformational uncertainties. Introduction of secondary sites usually requires more extensive molecular modeling for the specification of proper flexible tail positions.

Model parameters are based on an active site model and structural indexes calculated in Apex-3D's Computational Chemistry module. The calculated atomic properties are rounded off before use, based on an estimated parameter error. This helps to avoid chance correlations based on insignificant variability in the property. The parameters are divided into the following three groups:

1. Biophore site indexes

Charge, pi-population, electron donor index, electron acceptor index, HOMO, LUMO, atomic hydrophobicity, atomic refractivity

2. Secondary site indexes

H-acceptors (presence, pi-population, charge, electron donor, hydrophobicity, refractivity)

H-donors (presence, pi-population, charge, hydrophobicity, refractivity)

Heteroatoms (presence, pi-population, charge, electron donor, hydrophobicity, refractivity, formal charge)

Hydrophobic (presence, pi-population, charge, electron donor, hydrophobicity, refractivity)

Steric (presence, pi-population, charge, electron donor, hydrophobicity, refractivity, formal charge) Ring centers (presence, size, number of pi-electrons)

3. Global molecular properties

Positions of the secondary sites are selected from the positions of atoms in molecular superimpositions. An atom of a molecule occupies the secondary site if its distance from the site position is less than the user-specified site radius. To select only the most reasonable secondary sites you can also specify the site occupancy threshold--the minimal number of compounds occupying a site before it can be included as a site .

Secondary sites serve three primary purposes:

1. Identify possible extensions of the biophore common to the compounds in the model, for example, a region of space relative to the biophore with additional hydrogen-bond interactions which *increase* activity.

2. Identify steric interference; regions of space which when occupied by the ligand *decrease* activity.

3. Identify hydrophobic pockets; regions of space which when occupied by the ligand *increase* activity.

The biophore chosen for 3D-QSAR model building serves as a reference for superimposing the ligands. Biophore sites may also contribute quantitatively to the 3D-QSAR model as additional parameters.

4.7 3D-QSAR MODEL BUILDING PROCEDURE

1. Automated selection of biophores.

2. Optimization of superimposition of compounds sharing a common biophore.

3. Interactive specification of the 3D-QSAR model parameters based on physicochemical properties of biophoric features, secondary sites, and global molecular properties.

4. Calculation of the best 3D-QSAR model for the selected biophores.

5. Selection of the multiple regression equation using stepwise multiple regression.

6. Estimation of non-randomness and predictive power of obtained models and filtering out unreliable models. An example of a 3D-QSAR model for a small set of angiotensin-converting enzyme inhibitors is presented.

4.7.1 SELECTING RELIABLE 3D-QSAR MODELS

The basic statistical tool for 3D-QSAR model building in Apex-3D is the stepwise regression algorithm (Myers 1990). This algorithm selects multiple regression equations by deleting and adding variables using the partial F-test criterion. Variables are added only if they increase the predictability of the model based on the PRESS (PREdicted Sum of Squares.

4.8 THE CATALYST METHODOLOGY

4.8.1 MOLECULE PREPARATION

The molecules are edited with the Catalyst 2D/ 3D sketch facility and optimized applying the CHARMm forcefield . The molecules are then subjected to conformational analysis using the Poling algorithm through the conFirm module. It has different user control to suit the particular kind of molecules.

4.8.2 HYPOTHESIS GENERATION

Catalyst allows the building of a model or a hypothesis distinguishing chemical features essential for a particular activity in a class of compounds in different ways viz.

(1) By generating a hypothesis automatically from a training set of molecules.

(2) By constructing a hypothesis manually, where the substructures and chemical functions are assembled under the specified geometric constraints between them.

(3) By converting a molecule to hypothesis.

(4) By using a template molecule.

During the automated generation of hypothesis the user has several control parameters (Spacing MinPoints, MinSubsetPoints, Superimposition Error, Misses, Feature Misses, Complete Misses, Tolerance Factor, CheakSuperPostion, Weight Variation, Mapping Coeff, Mcm, IdealHbondGeocemOnly, Variable Weight, Variable Tolerance) under the more hypothesis option in the Catalyst Hypogen Module.

The Goal is to find a chemical feature - based model that is predictive beyond the training set molecules . Because the number of variables inherent to problem, of this kind is large, many simplifying approximation are necessary in order to achieve a practical result.

4.8.3 VALIDATE HYPOTHESIS

The final step is the validation of the hypothesis, which can be accomplished by evaluating the statistical significance, scrambling activities with structures testing the predictive ability on a test set.

The output from a Catalyst hypothesis generation job is the ten lowest cost hypotheses found during the analysis that are different from each other. It is often not possible to discriminate between these by any simple statistical procedure particularly if the cost differences are small (less than 10 bits). Therefore, visual evaluation procedure at this point in the experiment is used.

4.8.4 APPROACH

A set of 24 molecules was selected, according to picking rules as the training set. For the test sets two separately service were taken the first series had 14 molecules & the other series had 12 molecules. Total 26 molecules were chosen for test set predictions. The relative binding affinity data for estrogen receptor legends was collected from literature. The 3D structures were built interactively using Catalyst version 4.5 and minimized using CHARMm forcefield. Diverse conformations which accessed conformational space defined within 10Kcal of the estimated global minimum, were generated for each compound using the poling algorithm.

SECTION-B

RESULTS AND DISCUSSION

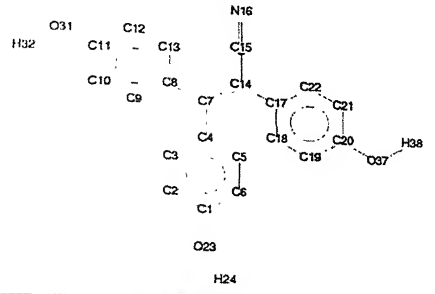
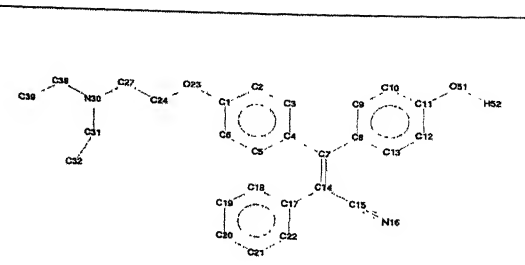
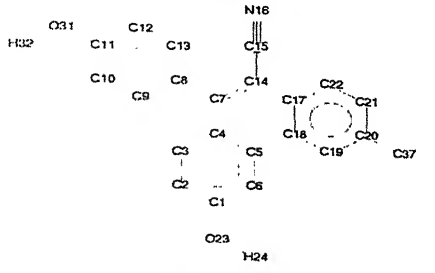
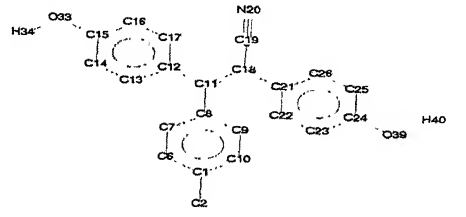
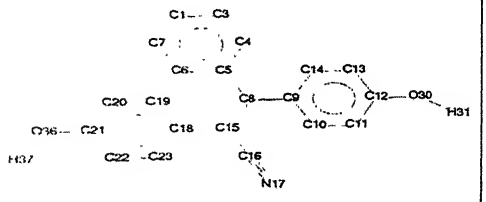
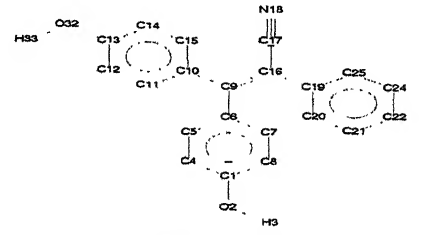
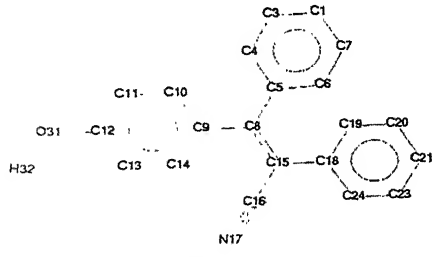
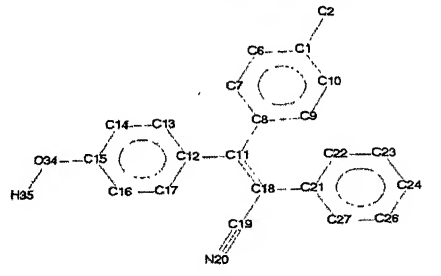
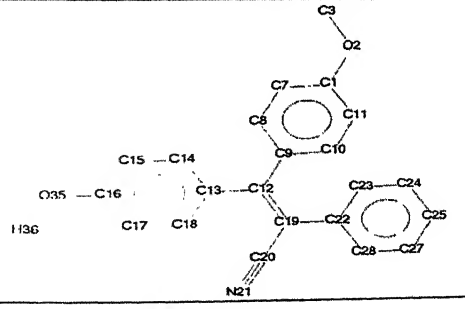
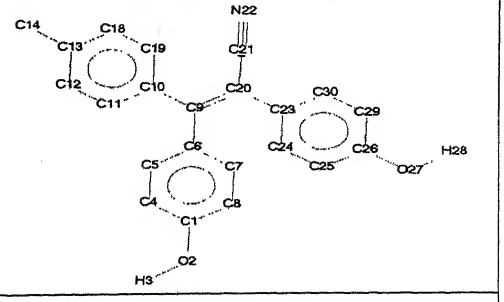
Out of several different estrogen antagonists reported only the 3,3,2-Triphenylacrylonitriles[182] **(table-1)** were considered for molecular modeling and 3D QSAR analysis by Catalyst and APEX-3D expert system running on an SGI Indy 4000 work-station. Since this series exhibits the largest range (over five (5) orders of magnitude) of RBA activities. Two different test sets (one with 1,1-Diphenylethylenes[183] and 2,3-Diarylindines[184, 185] were used assess the predictive ability of the 3D QSAR models generated.

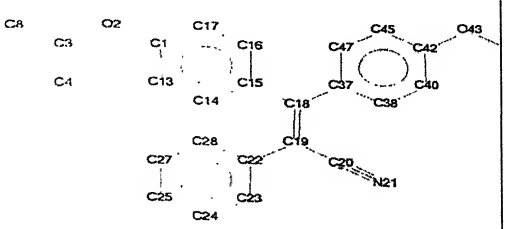
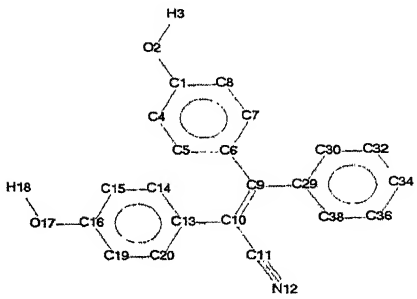
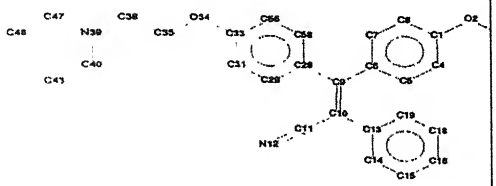
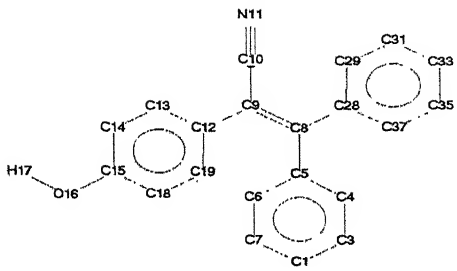
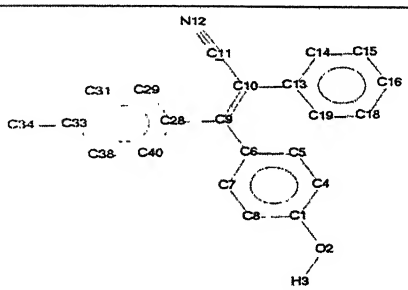
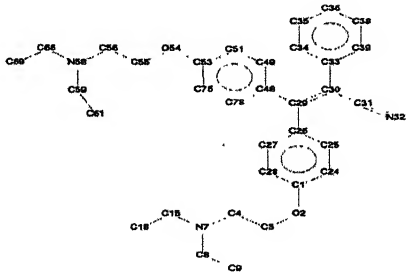
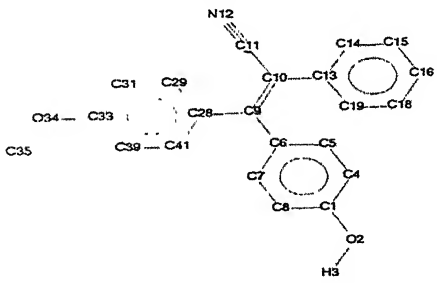
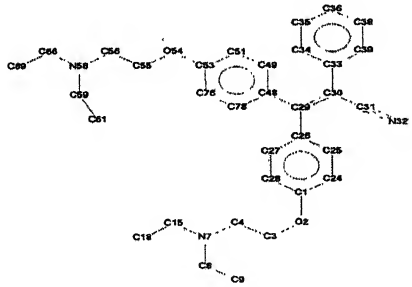
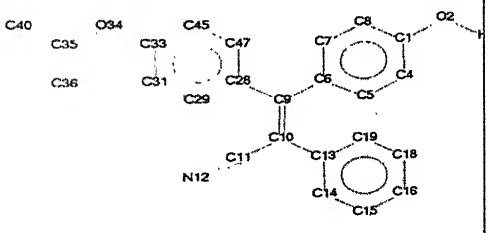
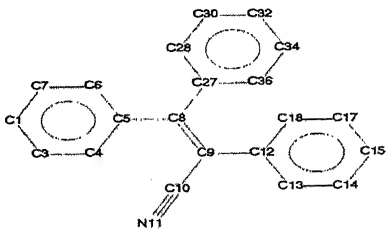
Common feature hypothesis were generated for finding the chemical features, in a 3D dimensional space, shared by a set of potent antiestrogenic molecules and the crystal structure of 4-hydroxy tamoxifen in its enzyme bound confirmation. The compound was extracted from 1ert of the protein data bank. 4-hydroxy tamoxifen was taken as the template on to which all the molecules were forced to map (principle = 2) and 10 of the most active molecules were as also considered in generating the hypothesis (principle value = 1).

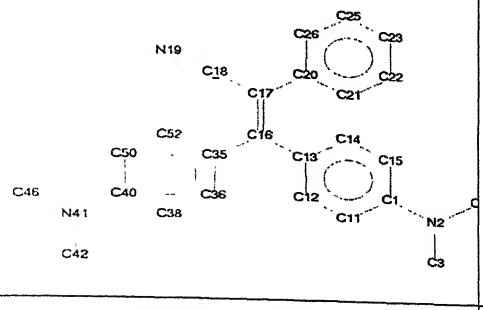
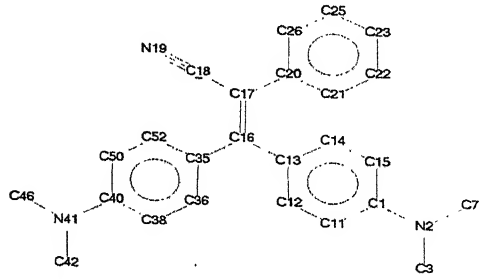
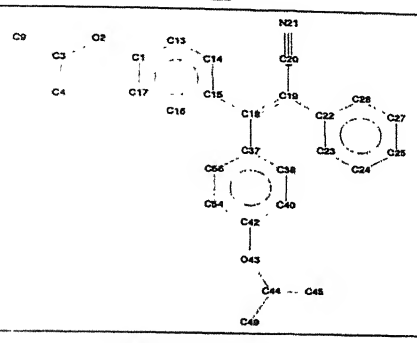
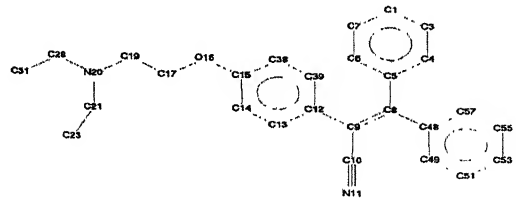
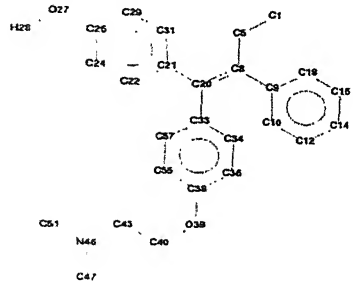
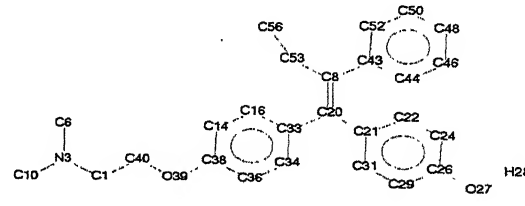
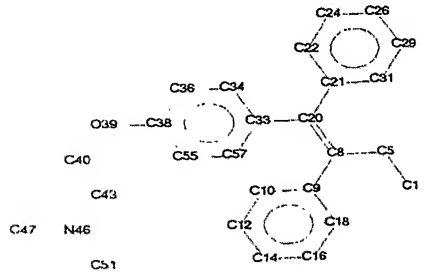
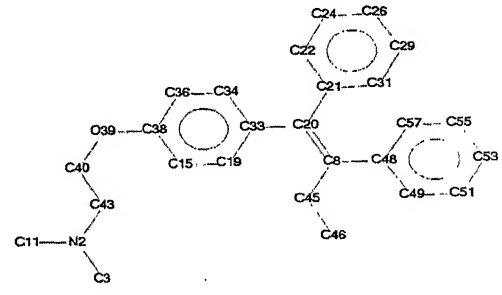
The pharamcophoric hypothesis represents the structural and function criteria required for estrogenic activity. The crystal structures of beta-esterdiol and Raloxifen in their enzyme bound conformation map well to the hypothesis further validating the pharmacophore.

All the molecules of the training set and test were aligned onto the hypothesis and the conformers were exported to APEX-3D for advanced 3D QSAR analysis.

TABLE-1: Molecules of the Training Set

	
JMC89_1	JMC89_2
	
JMC89_3	JMC89_4
	
JMC89_5	JMC89_6
	
JMC89_7	JMC89_8
	
JMC89_9	JMC89_10

	
JMC89_11	JMC89_12
	
JMC89_13	JMC89_14
	
JMC89_15	JMC89_16
	
JMC89_17	JMC89_18
	
JMC89_19	JMC89_20

	
JMC89_21	JMC89_22
	
JMC89_23	JMC89_24
	
JMC89_25	JMC89_26
	
JMC89_27	JMC89_28

MATERIAL AND METHODS

The molecules were stored in MDL format and were used for the computational calculation of different physicochemical properties including atomic charges, π-population, electron donor and acceptor indexes, HOMO and LUMO coefficient, hydrophobicity and molar refractivity based on atomic contributions by the MOPAC 6.0 (MNDO Hamiltonian)[186] version. The compounds were classified into following three classes (i) most active (log RBA>1.70) (ii) active (log RBA< 1.7) (iii) less active ($logIC_{50}$<1.0). The data were used by APEX-3D programme for automated biophore (pharmacophore) identification and 3D-QSAR model building. The automatically identified biophore (pharmacophore) by APEX-3D in terms of structural and electronic pattern, the local array of descriptor centres (like user defined atoms, pseudo atoms like ring centres, hydrophobic regions or hydrogen binding sites) which are common to a class of molecules in their bioactive conformation, responsible for activity through interaction with the receptor were used to derive 3D-QSAR equations with the setting of, the site radius at 1.20, the occupancy at 8, the sensitivity at 1.0 and the randomization at 100. The global properties, (total hydrophobicity and total refractivity) the biophoric site properties (π-population, charge, HOMO, LUMO, hydrogen acceptor, hydrogen donor, and hydrophobicity) and the secondary site parameters (hydrogen acceptor, presence; hydrogen donor, presence; heteroatom, presence; hydrophobic, hydrophobicity; steric, refractivity; ring, presence) were used as independent variables and biological activity as dependent variable, to derive equations for 3D-QSAR models[187-188].

Quality of each model was estimated from the observed R^2 (correlation coefficient between experimentally observed and calculated (APEX-3D uses the word approximated activity), RMSA (calculated root mean square error based on all compounds with degrees of freedom

correction), RMSP (root mean square error based on 'leave one out' with no degrees of freedom correction), chance statistics (evaluated as the ratio of the equivalent regression equation to the total number of randomized sets; a chance value of 0.1 corresponds to 10% chance of fortuitous correlation) and match parameter (evaluated for the quality of superimposition for molecules having common biophores; a value of 1 corresponding to the best possible fit 100%).

Several biophoric models were obtained with different sizes (number of biophoric sites) and arrangement (spatial orientation). Among several biophoric models, only ten (10) models were considered based on the statistical criteria ($R^2 > 0.65$, Chance<0.1, Superimposition match>0.7, RMSA< 0.9 and RMSP<1.0)[189] (table-2).

TABLE 2

STATISTICAL DETAILS OF THE BEST-SELECTED MODELS (MODEL NO.1-10)

S.N.	No	RMSA	RMSP	R2	Chance	Size	Match	Variable	No. of compounds
1	57	0.59	0.7	0.83	0	5	0.43	4	23
2	13	0.59	0.62	0.82	0	6	0.48	4	24
3	78	0.73	0.79	0.76	0	4	0.44	5	28
4	12	0.73	0.97	0.75	0.1	7	0.29	5	23
5	59	0.76	0.89	0.73	0.1	5	0.32	5	28
6	49	0.73	0.8	0.72	0.1	6	0.26	4	24
7	55	0.77	0.85	0.72	0.1	3	0.57	5	26
8	77	0.83	1.03	0.68	0.1	3	0.68	4	24
9	56	0.85	0.95	0.67	0.1	2	0.42	5	28
10	61	0.87	1.02	0.66	0.1	4	0.34	5	28

From the table it is clear that all the models 1-10 have high statistical significance >99%, according to the chance and F values. Model number 8 was rejected because it did not include the most active molecule (4-hydroxy tamoxifen).

MODEL-1

The five biophoric sites (1-5, white circles), common to all molecules in model no.1 correspond to the carbonyl oxygen, its lone pair and centre of the phenyl ring bearing the R_1 substituents respectively. The spatial disposition of these sites, in terms of inter site distances, is BS1-BS2=3.00 (±0.001), BS1-BS3=3.74 (±0.02), BS2-BS3=6.30 (±0.40) A^0 for model no. 1. The physicochemical characteristics of the biophore centres corresponding to sites are BS1: Pi-Popul [1.839±0.498], Charge_Het [-0.113±0.053], Don_01 [9.024±0.302], BS2: Pi-Popul [1.839±0.498], Charge_Het [-0.113±0.053], Don_01[9.024±0.302], BS3: H-Site [1±0], BS4: H-Site [1±0] and BS5: Cycle_size [1±0], Pi-electron [1±0].

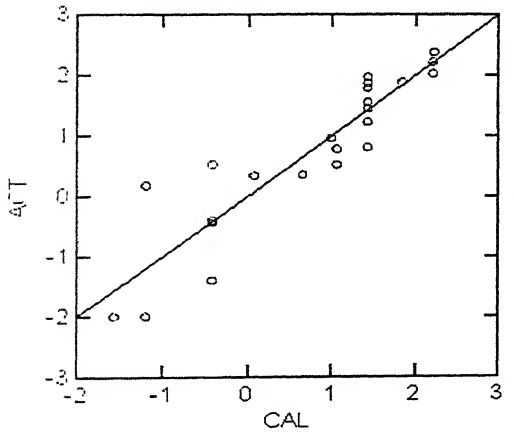
In addition to the identification of the five common key structural features described above as biophoric sites common to twenty three molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to four secondary site parameters (variables): **HYDROPHOBICITY** as a global (whole molecule) property, **steric[REFRACTIVITY]** at various sites as shown below.

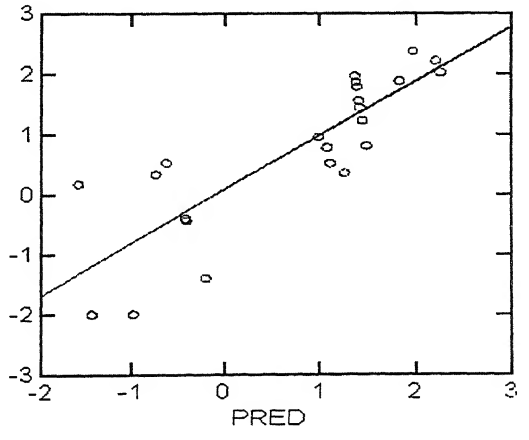
log (RBA) = 10.517(±1.174) HYDROPHOBICITY - 0.227 (±0.083) Steric [REFRACTIVITY] -0.823(0.149) Steric [REFRACTIVITY] + 0.454 (±0.133) Steric [REFRACTIVITY] - .980

This model presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

TABLE 3

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	0.99	0.99
JMC89_11	0.81	1.43	1.49
JMC89_12	0.78	1.06	1.08
JMC89_13	0.53	-0.42	-0.63
JMC89_14	0.52	1.06	1.11
JMC89_16	0.34	0.07	-0.74
JMC89_17	0.18	-1.2	-1.58
JMC89_18	-0.4	-0.42	-0.42
JMC89_19	-0.44	-0.42	-0.41
JMC89_1	2.22	2.21	2.21
JMC89_22	-1.4	-0.42	-0.2
JMC89_23	-2	-1.2	-0.98
JMC89_24	-2	-1.57	-1.43





MODEL-2

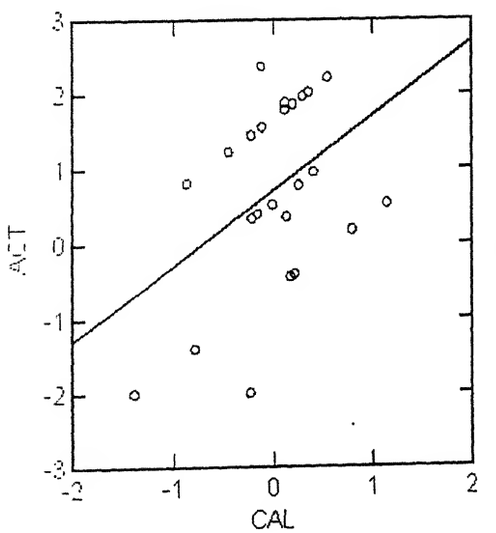
In addition to the identification of the five common key structural features described above as biophoric sites common to twenty-four molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Eestrogen receptor binding activity was related to four secondary site parameters .

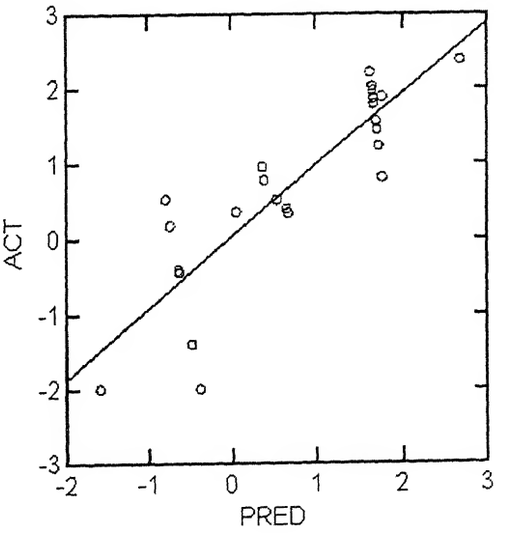
log (RBA) = -2.994 (±0.671) PI-POPUL + 9.161 (±1.055) HYDROPHOBICITY + 7.693 (±2.457) Hydrophobic [HYDROPHOBICITY] + 0.325 (±0.110) Steric [REFRACTIVITY] + 0.697

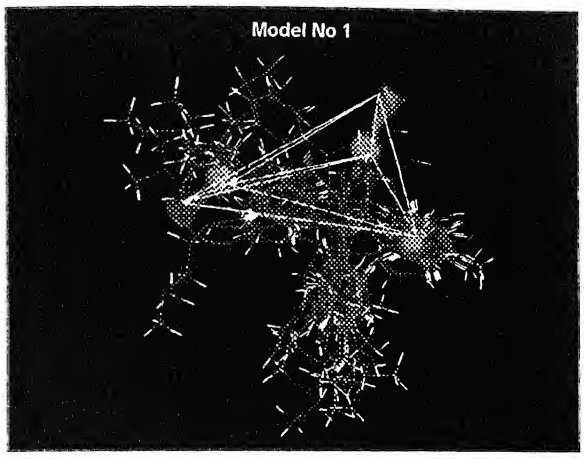
This model also presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

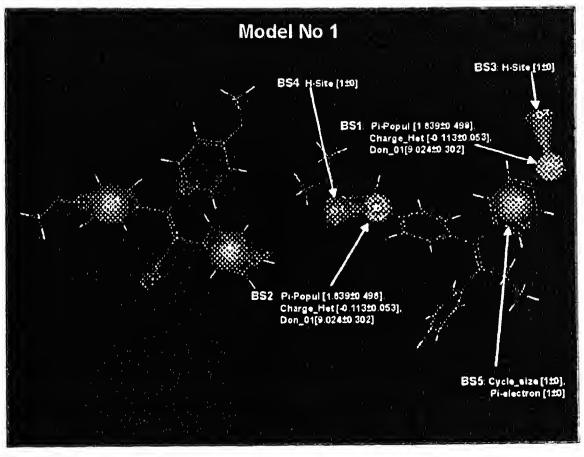
TABLE 4

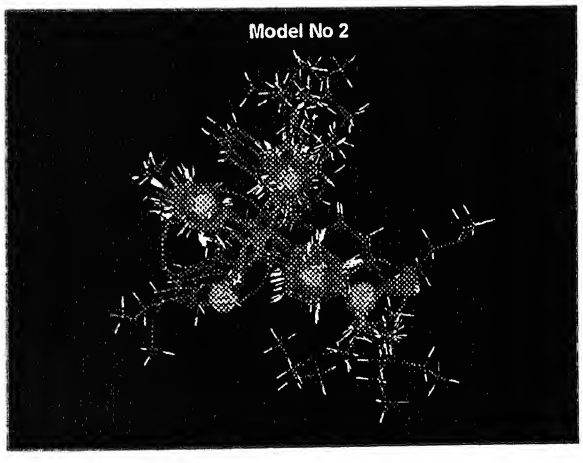
Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	0.4	0.37
JMC89_11	0.81	-0.87	1.77
JMC89_12	0.78	0.25	0.39
JMC89_13	0.53	1.14	-0.79
JMC89_14	0.52	-0.01	0.54
JMC89_15	0.4	-0.16	0.65
JMC89_16	0.34	-0.22	0.67
JMC89_17	0.18	0.79	-0.74
JMC89_18	-0.4	0.21	-0.64
JMC89_19	-0.44	0.17	-0.63
JMC89_1	2.22	0.54	1.63
JMC89_22	-1.4	-0.79	-0.48
JMC89_23	-2	-1.39	-0.38
JMC89_24	-2	-0.24	-1.58
JMC89_2	2.03	0.35	1.65
JMC89_3	1.97	0.29	1.66
JMC89_4	1.89	0.11	1.77
JMC89_5	1.87	0.19	1.67
JMC89_6	1.79	0.11	1.67
JMC89_7	1.56	-0.12	1.7
JMC89_8	1.45	-0.23	1.71
JMC89_9	1.23	-0.45	1.73
TAM_OHE1	0.36	0.13	0.06
TAM_PDB	2.38	-0.13	2.68

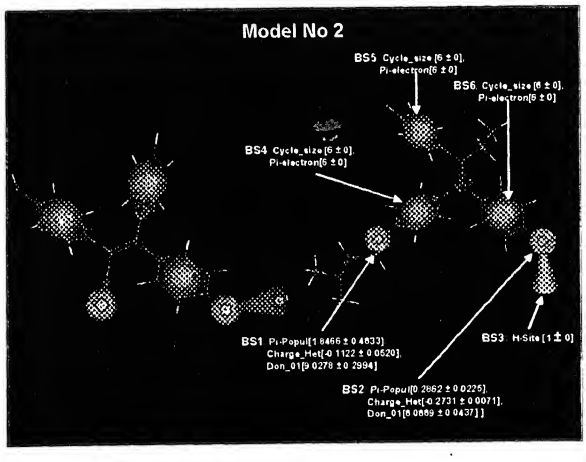





Model No 1


Model No 1
BS3 H-Site [1±0]
BS4 H-Site [1±0]
BS1 Pi-Popul [1.839±0.498],
Charge_Het [-0.113±0.053],
Don_01[9.024±0.302]
BS2 Pi-Popul [1.839±0.498],
Charge_Het [-0.113±0.053],
Don_01[9.024±0.302]
BS5 Cycle_size [1±0],
Pi-electron [1±0]


Model No 2


Model No 2
BS5 Cycle_size [6 ± 0],
Pi-electron[6 ± 0]
BS6 Cycle_size [6 ± 0],
Pi-electron[6 ± 0]
BS4 Cycle_size [6 ± 0],
Pi-electron[6 ± 0]
BS1 Pi-Popul[1.8466 ± 0.4633],
Charge_Het[-0.1122 ± 0.0520],
Don_01[9.0278 ± 0.2994]
BS3 H-Site [1 ± 0]
BS2 Pi-Popul[0.2662 ± 0.0225],
Charge_Het[-0.2731 ± 0.0071],
Don_01[6.0669 ± 0.0437]]

MODEL-3

The physicochemical properties of the biophoric centers corresponding to sites BS1: Don_01[8.163 ± 0.254], BS2: Cycle_size [6±0], Pi-electron [6±0], BS3: Cycle_size [6±0], Pi-electron [6±0], BS4: Cycle_size [6±0], Pi-electron [6±0].

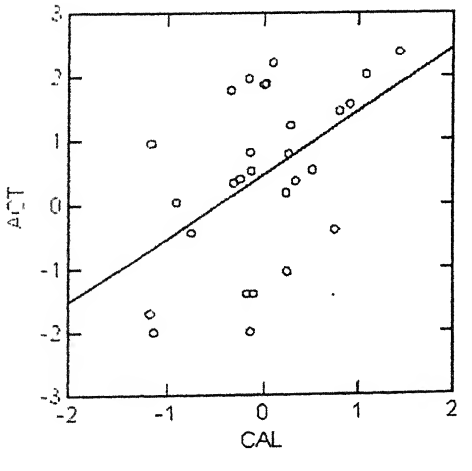
In addition to the identification of the four common key structural features described above as biophoric sites common to twenty eight molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Eestrogen receptor binding activity was related to four secondary site parameters.

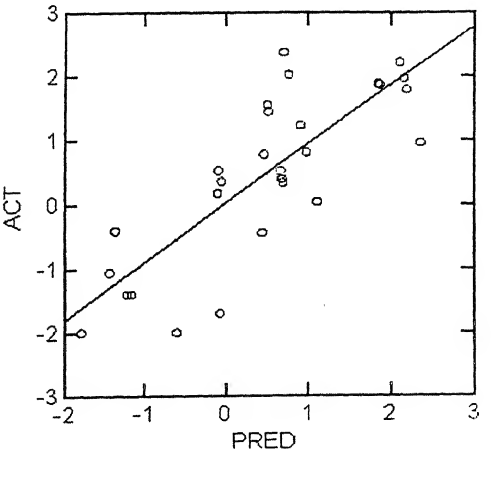
log (RBA) = 4.685 (±1.332) HYDROPHOBICITY -8.13 (±3.899) Hydrophobic [HYDROPHOBICITY] - 0.867 (±0.238) Steric [REFRACTIVITY] -0.724 (±0.129) Steric [REFRACTIVITY] -0.428 (±0.106) Steric [REFRACTIVITY]

The model also presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

TABLE 5

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	-1.17	2.36
JMC89_11	0.81	-0.14	0.98
JMC89_12	0.78	0.26	0.47
JMC89_13	0.53	0.51	-0.09
JMC89_14	0.52	-0.13	0.67
JMC89_15	0.4	-0.25	0.69
JMC89_16	0.34	-0.31	0.7
JMC89_17	0.18	0.23	-0.1
JMC89_18	-0.4	0.75	-1.36
JMC89_19	-0.44	-0.75	0.45
JMC89_1	2.22	0.09	2.11
JMC89_20	-1.05	0.24	-1.43
JMC89_21	-1.4	-0.11	-1.21
JMC89_22	-1.4	-0.18	-1.16
JMC89_23	-2	-1.14	-0.61
JMC89_24	-2	-0.14	-1.78
JMC89_2	2.03	1.08	0.77
JMC89_3	1.97	-0.16	2.16
JMC89_4	1.89	0.02	1.85
JMC89_5	1.87	0	1.87
JMC89_6	1.79	-0.34	2.19
JMC89_7	1.56	0.91	0.51
JMC89_8	1.45	0.8	0.52
JMC89_9	1.23	0.28	0.91
TAM_E1	-1.7	-1.18	-0.07
TAM_OHE1	0.36	0.33	-0.05
TAM_PDB	2.38	1.43	0.71
TAM_Z1	0.04	-0.91	1.11





MODEL-4

The physicochemical properties of the biophoric centers corresponding to sites BS1: Pi-Popul [1.915±0.356], Charge_Het [-0.293±1.912], Don_01[9.072±0.211], BS2: Pi-Popul [0.288± 0.022], Charge_Het 0.091± 1.746], Don_01[8.107±0.048], BS3: H-Site[1 ±0], BS4: H-Site[1 ±0], BS5: Cycle_size [1±0], Pi-electron [1±0], BS6: Cycle_size [1±0], Pi-electron [1±0], BS7: Cycle_size [1±0], Pi-electron [1±0].

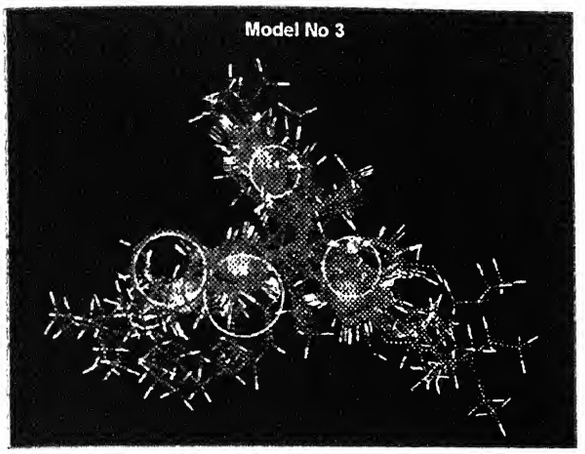
In addition to the identification of the four common key structural features described above as biophoric sites common to twenty three molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Eestrogen receptor binding activity was related to four secondary site parameters.

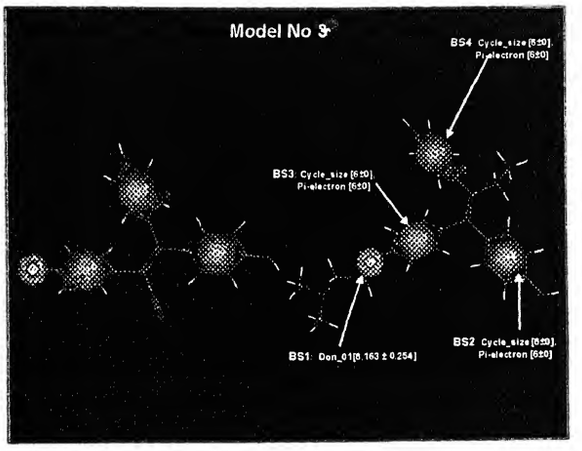
log (RBA) = 3.625 (±1.455) HYDROPHOBICITY + 0.734 (±0.232) Steric [REFRACTIVITY] -0.248(±0.105) Steric [REFRACTIVITY] -0.53(±0.167) Steric [REFRACTIVITY] + 0.297 (±0.09) Steric [REFRACTIVITY].

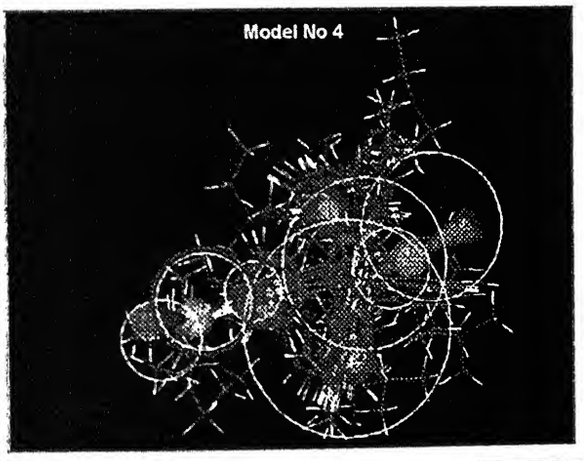
The model also presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

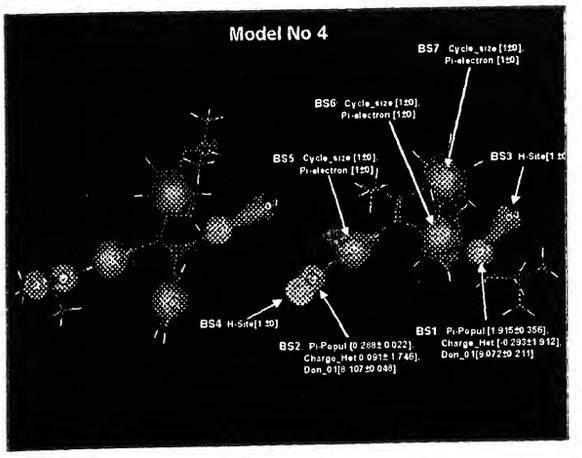
TABLE 6

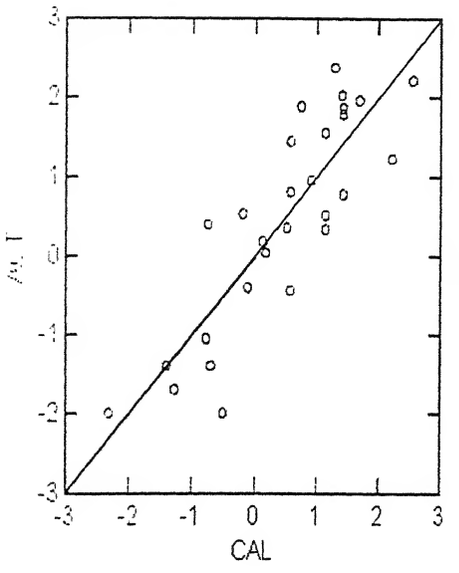
Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	-0.78	1.93
JMC89_11	0.81	0.29	0.39
JMC89_12	0.78	0.22	0.52
JMC89_13	0.53	-0.03	0.57
JMC89_14	0.52	-0.9	1.55
JMC89_15	0.4	1.4	-2.14
JMC89_16	0.34	-0.22	0.61
JMC89_17	0.18	0.03	0.12
JMC89_18	-0.4	-0.06	-0.32
JMC89_19	-0.44	-0.93	0.72
JMC89_1	2.22	0.48	1.62
JMC89_22	-1.4	-0.34	-0.93
JMC89_23	-2	-1.51	0.4
JMC89_24	-2	0.02	-3.34
JMC89_2	2.03	1.2	0.42
JMC89_3	1.97	0.23	1.68
JMC89_4	1.89	0.25	1.6
JMC89_5	1.87	0.23	1.6
JMC89_6	1.79	0.37	1.37
JMC89_7	1.56	-0.08	1.66
JMC89_13	0.53	-0.2	-0.43
JMC89_14	0.52	1.14	1.19
JMC89_15	0.4	-0.76	-1.04
JMC89_16	0.34	1.14	1.21
JMC89_17	0.18	0.12	0.09
JMC89_18	-0.4	-0.12	-0.04
JMC89_19	-0.44	0.57	0.65
JMC89_1	2.22	2.55	2.68
JMC89_20	-1.05	-0.78	-0.57
JMC89_21	-1.4	-1.4	-1.41
JMC89_22	-1.4	-0.71	-0.2
JMC89_23	-2	-0.51	0.32
JMC89_24	-2	-2.31	-2.7
JMC89_2	2.03	1.41	1.28
JMC89_3	1.97	1.69	1.64
JMC89_4	1.89	0.74	0.66
JMC89_5	1.87	1.43	1.37
JMC89_6	1.79	1.43	1.38
JMC89_7	1.56	1.14	1.1
JMC89_8	1.45	0.57	0.5
JMC89_9	1.23	2.21	2.46
TAM_E1	-1.7	-1.28	-0.57
TAM_OHE1	0.36	0.51	0.53
TAM_PDB	2.38	1.29	1.06
TAM_Z1	0.04	0.17	0.19

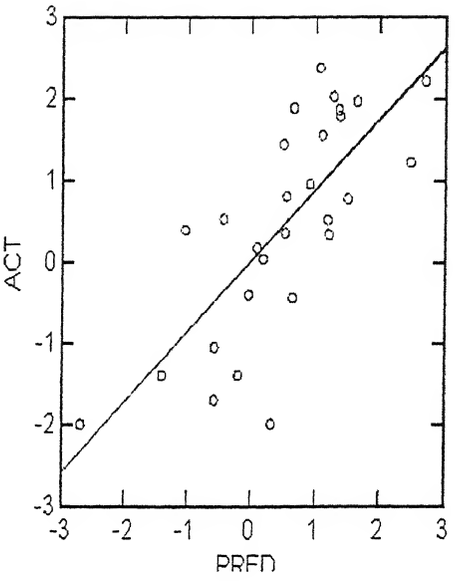

Model No 3


Model No 3
BS4 Cycle_size [6±0], Pi-electron [6±0]
BS3 Cycle_size [6±0], Pi-electron [6±0]
BS2 Cycle_size [6±0], Pi-electron [6±0]
BS1 Don_01[8.163 ± 0.254]


Model No 4


Model No 4
BS7 Cycle_size [1±0], Pi-electron [1±0]
BS6 Cycle_size [1±0], Pi-electron [1±0]
BS5 Cycle_size [1±0], Pi-electron [1±0]
BS3 H-Site[1 ±
BS4 H-Site[1 ±0]
BS2 Pi-Popul [0.268± 0.022], Charge_Het 0.091± 1.746], Don_01[8.107±0.048]
BS1 Pi-Popul [1.915±0.356], Charge_Het [-0.293±1.912], Don_01[9.072±0.211]





MODEL-5

The physicochemical properties of the biophoric centers corresponding to sites BS1: Don_01[8.166 ± 0.253], BS2: H-Site[1±0], BS3: Cycle_size [6±0], Pi-electron [6±0], BS4: Cycle_size [6±0], Pi-electron [6±0], BS5: Cycle_size [6±0], Pi-electron [6±0].

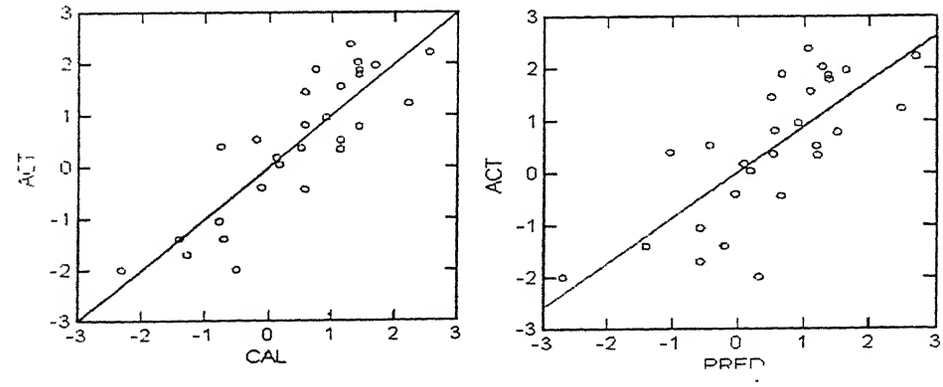
In addition to the identification of the seven common key structural features described above as biophoric sites common to twenty three molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to five secondary site parameters.

log (RBA) = -1.143 (±0.336) TOTAL_HYDROPHOBICITY+1.894 (±0.447) Hydrophobic [HYDROPHOBICITY]-7.789 (±3.42) Hydrophobic [HYDROPHOBICITY] + 0.372 (±0.164) Steric [REFRACTIVITY] - 0.408(±0.158)Steric [REFRACTIVITY].

The model also presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

TABLE-7

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	0.91	0.91
JMC89_11	0.81	0.57	0.55
JMC89_12	0.78	1.43	1.51
JMC89_13	0.53	-0.2	-0.43
JMC89_14	0.52	1.14	1.19
JMC89_15	0.4	-0.76	-1.04
JMC89_16	0.34	1.14	1.21
JMC89_17	0.18	0.12	0.09
JMC89_18	-0.4	-0.12	-0.04
JMC89_19	-0.44	0.57	0.65
JMC89_1	2.22	2.55	2.68
JMC89_20	-1.05	-0.78	-0.57
JMC89_21	-1.4	-1.4	-1.41
JMC89_22	-1.4	-0.71	-0.2
JMC89_23	-2	-0.51	0.32
JMC89_24	-2	-2.31	-2.7
JMC89_2	2.03	1.41	1.28
JMC89_3	1.97	1.69	1.64
JMC89_4	1.89	0.74	0.66
JMC89_5	1.87	1.43	1.37
JMC89_6	1.79	1.43	1.38
JMC89_7	1.56	1.14	1.1
JMC89_8	1.45	0.57	0.5
JMC89_9	1.23	2.21	2.46
TAM_E1	-1.7	-1.28	-0.57
TAM_OHE1	0.36	0.51	0.53
TAM_PDB	2.38	1.29	1.06
TAM_Z1	0.04	0.17	0.19



MODEL-6

The physicochemical properties of the biophoric centers corresponding to sites BS1: Pi-Popul [0.2889± 0.0202], Charge_Het [-0.2744±0.0048], Don_01[8.1019±0.0510], BS2: Pi-Popul [1.7684± 0.5971], Charge_Het [1.7684±0.5971], Don_01[8.954±0.4521], BS3: H-Site[1±0], BS4: H-Site[1±0], BS5: Cycle_size [6±0], Pi-electron [6±0], BS6: Cycle_size [6±0], Pi-electron [6±0].

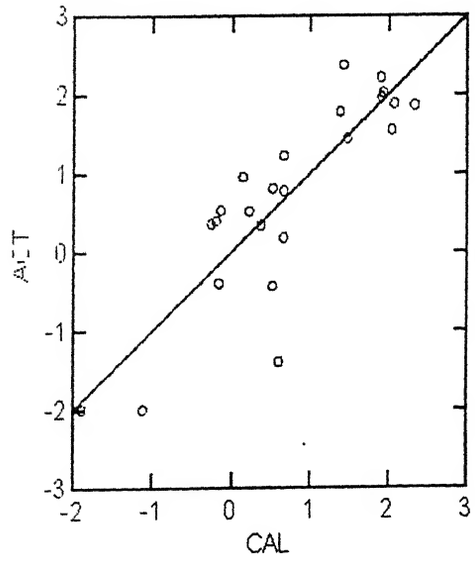
In addition to the identification of the seven common key structural features described above as biophoric sites common to all twenty eight molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to four secondary site parameters.

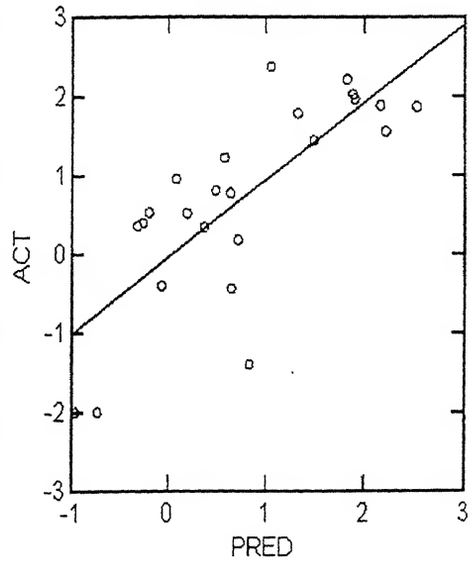
Log RBA = -1.142 (±0.334) TOTAL_HYDROPHOBICITY + 4.841 (±1.118) Hydrophobic [HYDROPHOBICITY] -0.485 (±0.118) Steric [REFRACTIVITY] + 0.601(±0.143) Steric [REFRACTIVITY] + 7.813

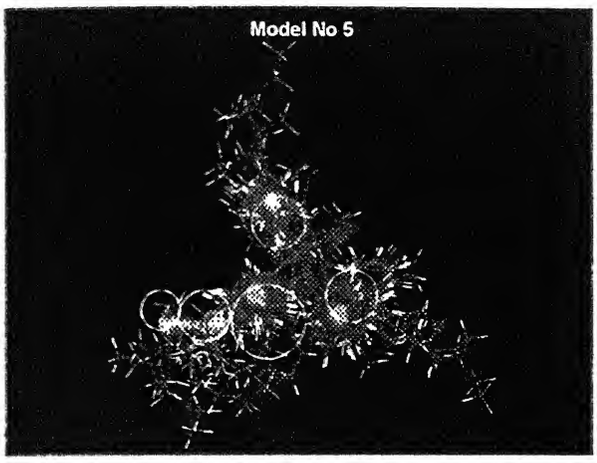
This model too presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

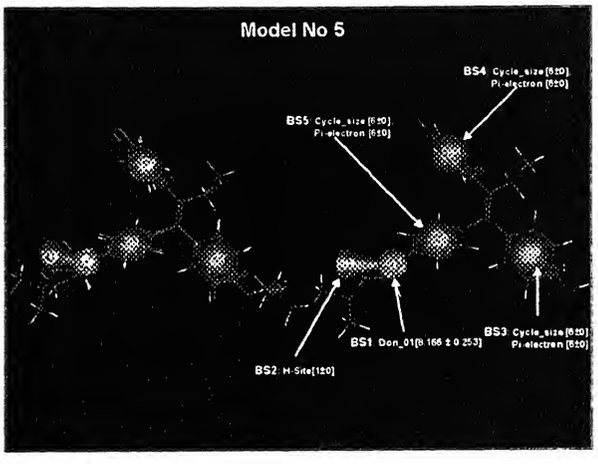
TABLE-8

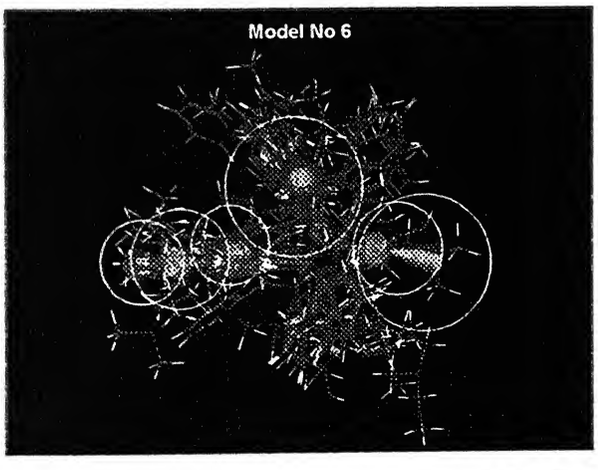
Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	0.14	0.08
JMC89_11	0.81	0.52	0.49
JMC89_12	0.78	0.66	0.64
JMC89_13	0.53	-0.14	-0.2
JMC89_14	0.52	0.22	0.19
JMC89_15	0.4	-0.2	-0.26
JMC89_16	0.34	0.37	0.37
JMC89_17	0.18	0.66	0.72
JMC89_18	-0.4	-0.16	-0.07
JMC89_19	-0.44	0.52	0.65
JMC89_1	2.22	1.9	1.82
JMC89_22	-1.4	0.6	0.84
JMC89_23	-2	-1.12	-0.73
JMC89_24	-2	-1.89	-0.96
JMC89_2	2.03	1.93	1.88
JMC89_3	1.97	1.91	1.9
JMC89_4	1.89	2.07	2.16
JMC89_5	1.87	2.33	2.52
JMC89_6	1.79	1.38	1.32
JMC89_7	1.56	2.04	2.21
JMC89_8	1.45	1.47	1.48
JMC89_9	1.23	0.66	0.58
TAM_OHE1	0.36	-0.26	-0.32
TAM_PDB	2.38	1.42	1.05

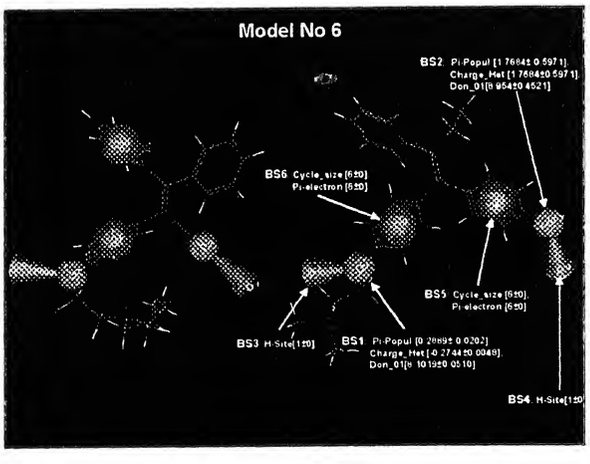





Model No 5


Model No 5
BS4: Cycle_size [6±0]
Pi-electron [6±0]
BS5: Cycle_size [6±0]
Pi-electron [6±0]
BS3: Cycle_size [6±0]
Pi-electron [6±0]
BS1: Don_01[8.166 ± 0.253]
BS2: H-Site[1±0]


Model No 6


Model No 6
BS2: Pi-Popul [1.76845 0.597 1]
Charge_Het [1.76845 0.597 1]
Don_01[8.95470 4521]
BS6: Cycle_size [6±0]
Pi-electron [6±0]
BS5: Cycle_size [6±0]
Pi-electron [6±0]
BS3: H-Site[1±0]
BS1: Pi-Popul [0.28995 0.0202]
Charge_Het [-0.27445 0.0048]
Don_01[8.10192 0.0510]
BS4: H-Site[1±0]

MODEL-7

The physicochemical properties of the biophoric centers corresponding to sites BS1: Pi-Popul [0.2917± 0.0219], Charge_Het [-0.2747±0.0062], Don_01[8.0934±0.0543], BS2: Cycle_size [6±0], Pi-electron [6±0], BS3: Cycle_size [6±0], Pi-electron [6±0].

In addition to the identification of the seven common key structural features described above as biophoric sites common to twenty seven molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to five secondary site parameters.

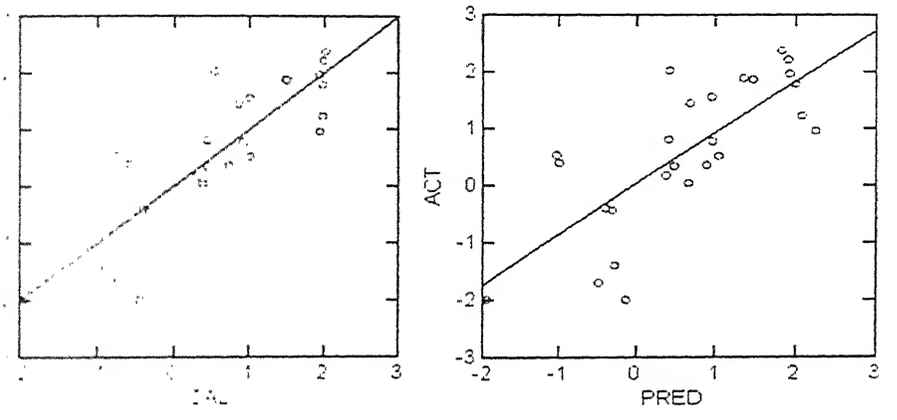
Log RBA = -2.053 (±0.372) TOTAL_HYDROPHOBICITY + 5.913(±2.497) Hydrophobic [HYDROPHOBICITY] + 1.516 (±0.417) Hydrophobic [HYDROPHOBICITY] -0.296 (±0.121) Steric [REFRACTIVITY] - 0.622 (±0.159) Steric [REFRACTIVITY] + 13.511

The model presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

TABLE-9

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	1.94	2.25
JMC89_11	0.81	0.44	0.41
JMC89_12	0.78	0.92	0.96
JMC89_13	0.53	-0.75	-1.03
JMC89_14	0.52	1	1.04
JMC89_15	0.4	-0.58	-1
JMC89_16	0.34	0.44	0.48
JMC89_17	0.18	0.31	0.38
JMC89_18	-0.4	-0.4	-0.4
JMC89_19	-0.44	-0.32	-0.31
JMC89_1	2.22	2	1.91
JMC89_22	-1.4	-0.9	-0.28
JMC89_23	-2	-1.96	-1.94
JMC89_24	-2	-0.44	-0.13
JMC89_2	2.03	0.54	0.42
JMC89_3	1.97	1.94	1.93

JMC89_4	1.89	1.48	1.35
JMC89_5	1.87	1.51	1.47
JMC89_6	1.79	1.98	2
JMC89_7	1.56	1	0.95
JMC89_8	1.45	0.86	0.68
JMC89_9	1.23	1.98	2.08
TAM_E1	-1.7	-0.75	-0.49
TAM_OHE1	0.36	0.73	0.89
TAM_PDB	2.38	2.02	1.83



MODEL-9

The physicochemical properties of the biophoric centers corresponding to sites BS1: Cycle_size [6±0], Pi-electron [6±0], BS2: Cycle_size [6±0], Pi-electron [6±0].

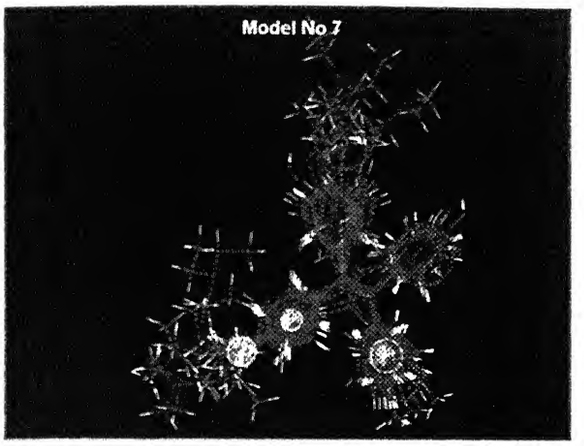
In addition to the identification of the seven common key structural features described above as biophoric sites common to twenty seven molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to five secondary site parameters.

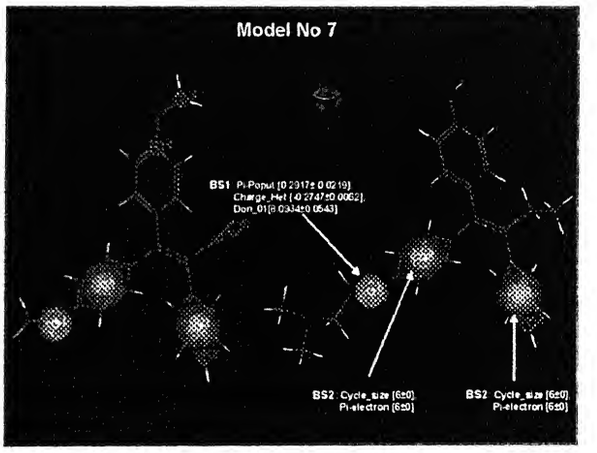
Log RBA = -0.743(±0.391) TOTAL_HYDROPHOBICITY - 5.679(±1.649) Hydrophobic [HYDROPHOBICITY] -0.524 (±0.14) Steric [REFRACTIVITY] + 5.563(±1.23) Steric [REFRACTIVITY]+4.284 (±1.607) Steric [REFRACTIVITY] -30.464.

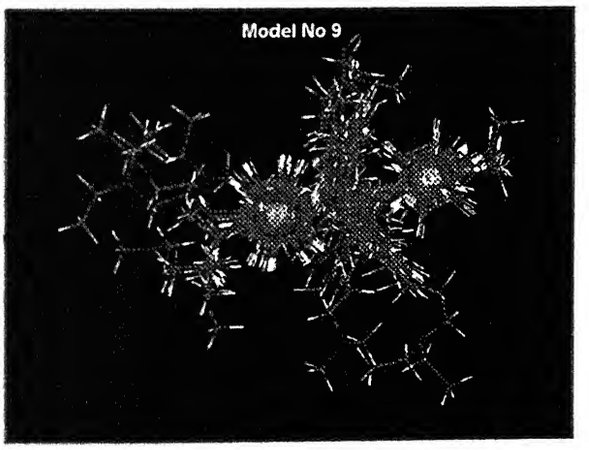
The model too presented good predictions (R^2 and LOO R^2) for the training set as shown in table and plot below.

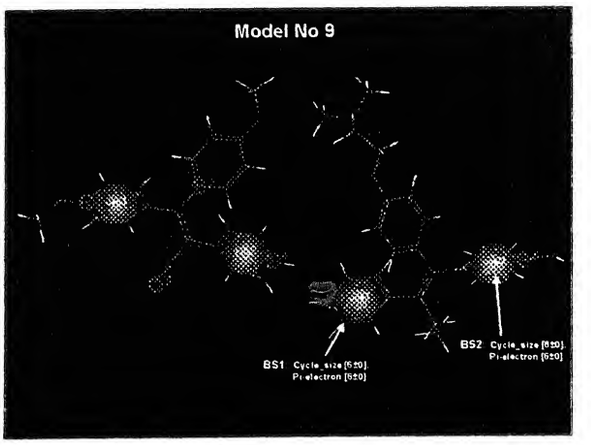
TABLE-10

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	1.59	1.83
JMC89_11	0.81	-0.18	-0.28
JMC89_12	0.78	1.22	1.29
JMC89_13	0.53	0.7	0.72
JMC89_14	0.52	1.04	1.09
JMC89_15	0.4	0.47	0.5
JMC89_16	0.34	-0.25	-0.48
JMC89_17	0.18	1.61	2.11
JMC89_18	-0.4	0.22	0.41
JMC89_19	-0.44	-0.18	-0.16
JMC89_1	2.22	3.12	3.44
JMC89_20	-1.05	-0.47	-0.28
JMC89_21	-1.4	-0.45	-0.3
JMC89_22	-1.4	-1.08	-0.85
JMC89_23	-2	-2.19	-2.29
JMC89_24	-2	-1.29	-1.08
JMC89_2	2.03	0.7	0.58
JMC89_3	1.97	1.59	1.44
JMC89_4	1.89	1.71	1.67
JMC89_5	1.87	0.63	0.23
JMC89_6	1.79	1.22	1.14
JMC89_7	1.56	1.04	0.98
JMC89_8	1.45	0.67	0.59
JMC89_9	1.23	1.22	1.22
TAM_E1	-1.7	-0.45	-0.25
TAM_OHE1	0.36	-0.22	-0.28
TAM_PDB	2.38	1.01	0.47
TAM_Z1	0.04	-0.08	-0.5
JMC89_16	0.34	-0.25	-0.48
JMC89_17	0.18	1.61	2.11
JMC89_18	-0.4	0.22	0.41
JMC89_19	-0.44	-0.18	-0.16
JMC89_1	2.22	3.12	3.44
JMC89_20	-1.05	-0.47	-0.28
JMC89_21	-1.4	-0.45	-0.3
JMC89_22	-1.4	-1.08	-0.85
JMC89_23	-2	-2.19	-2.29
JMC89_24	-2	-1.29	-1.08
JMC89_2	2.03	0.7	0.58
JMC89_3	1.97	1.59	1.44
JMC89_4	1.89	1.71	1.67

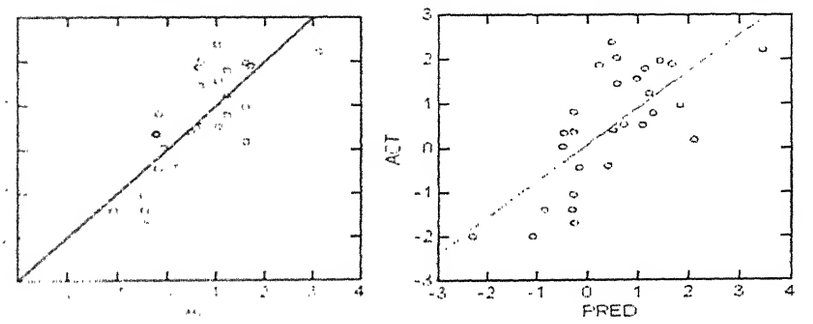

Model No 7


Model No 7
BS1 Pi-Popul (0.2917± 0.0219)
Charge_Het (-0.2747±0.0002)
Don_01(8.0334±0.0543)
BS2 Cycle_size (6±0)
Pi-electron (6±0)
BS2 Cycle_size (6±0)
Pi-electron (6±0)


Model No 9


Model No 9
BS2 Cycle_size (6±0)
Pi-electron (6±0)
BS1 Cycle_size (6±0)
Pi-electron (6±0)

JMC89_5	1.87	0.63	0.23
JMC89_6	1.79	1.22	1.14
JMC89_7	1.56	1.04	0.98
JMC89_8	1.45	0.67	0.59
JMC89_9	1.23	1.22	1.22
TAM_E1	-1.7	-0.45	-0.25
TAM_OHE1	0.36	-0.22	-0.28
TAM_PDB	2.38	1.01	0.47
TAM_Z1	0.04	-0.08	-0.5

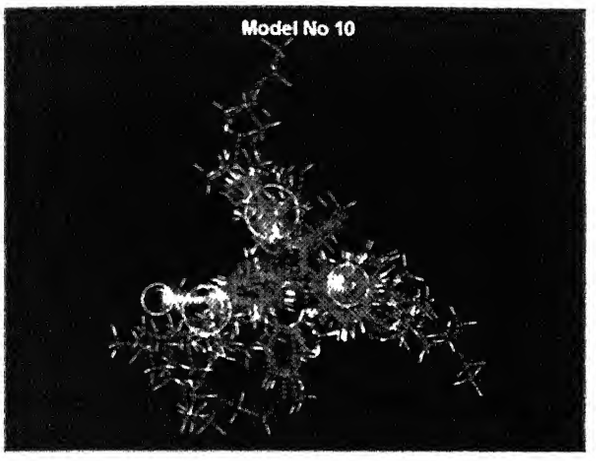


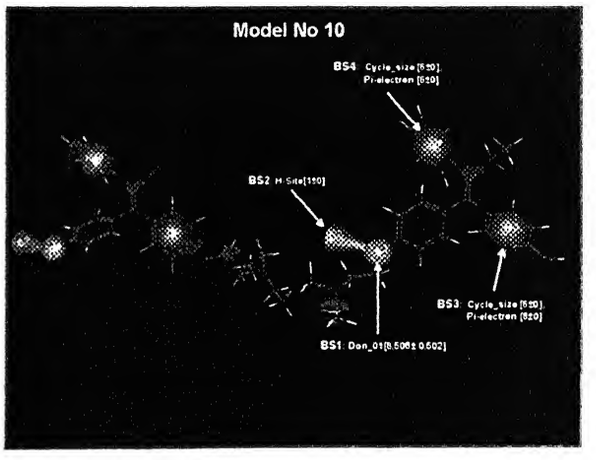
MODEL-10

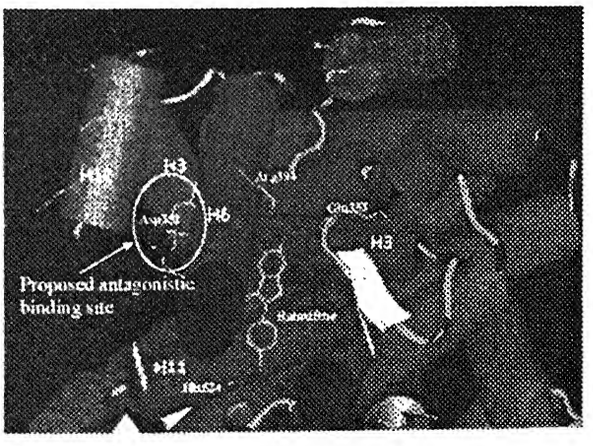
The physicochemical properties of the biophoric centers corresponding to sites BS1: Don_01[8.506± 0.502], BS2: H-Site[1±0], BS3: Cycle_size [6±0], Pi-electron [6±0], BS4: Cycle_size [6±0], Pi-electron [6±0].

In addition to the identification of the seven common key structural features described above as biophoric sites common to twenty seven molecules, three-dimensional multiparameter equations were derived using these pharmacophore as template for superimposition. The in vitro activity log (RBA) for the Estrogen receptor binding activity was related to five secondary site parameters.

log RBA = -1.596 (±0.367) TOTAL_HYDROPHOBICITY + 11.229 (±5.154) Hydrophobic [HYDROPHOBICITY] - 1.743 (± 0.898) Hydrophobic [HYDROPHOBICITY] -7.936 (± 2.012) Hydrophobic [HYDROPHOBICITY] + 1.5 (± 0.601) Hydrophobic [HYDROPHOBICITY] + 8.738


Model No 10


Model No 10
BS4: Cycle_size [550]
Pi-electron [550]
BS2: H-Site[110]
BS3: Cycle_size [550]
Pi-electron [520]
BS1: Don_01[5.5065 0.502]


H3
H6
H3
Proposed antagonistic
binding site

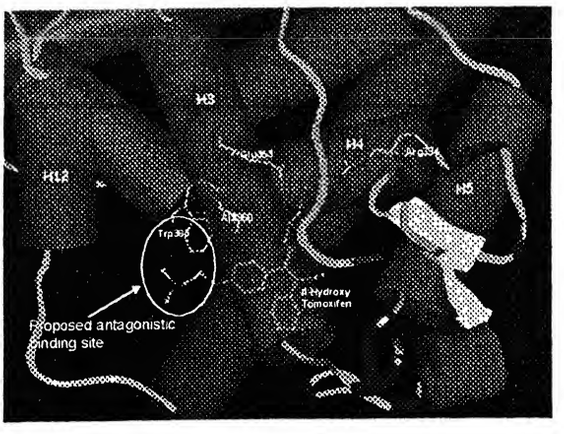

H3
H4
H12
H5
4-Hydroxy
Tamoxifen
Proposed antagonistic
binding site

TABLE-11

Molecule	log RBA	Calculated log RBA	Predicted log RBA
JMC89_10	0.96	0.36	0.33
JMC89_11	0.81	-0.12	-0.17
JMC89_12	0.78	1.08	1.11
JMC89_13	0.53	0.84	1.13
JMC89_14	0.52	0.68	0.69
JMC89_15	0.4	-1.43	-2.14
JMC89_16	0.34	0.68	0.7
JMC89_17	0.18	1.08	1.19
JMC89_18	-0.4	-1.08	-1.2
JMC89_19	-0.44	-0.12	-0.1
JMC89_1	2.22	2.59	3.02
JMC89_20	-1.05	0.2	0.27
JMC89_21	-1.4	-0.68	-0.6
JMC89_22	-1.4	-0.48	0.76
JMC89_23	-2	-1.4	-1.23
JMC89_24	-2	-0.44	-0.31
JMC89_2	2.03	1.94	1.91
JMC89_3	1.97	2.22	3.17
JMC89_4	1.89	1.58	1.3
JMC89_5	1.87	1.08	0.98
JMC89_6	1.79	1.08	0.99
JMC89_7	1.56	0.68	0.61
JMC89_8	1.45	1.1	0.77
JMC89_9	1.23	1.25	1.25
TAM_E1	-1.7	-0.68	-0.57
TAM_OHE1	0.36	-0.2	-0.23
TAM_PDB	2.38	1.78	1.52
TAM_Z1	0.04	-0.68	-0.76

VALIDATION OF THE MODELS

The models generated were validated for their predictive ability against a test set: two series of compounds were used as a test set, which were not the part of the training set.

Predictions of test set of 26 compounds by the hypotheses 1-10 derived by using the training set of 28 compounds.

TABLE-12

Test molecule	ACT	B57	B13	B78	B12	B59	B49	B55	B56	B61
1	1.58	0.39	7.34	2.13	1.42	0.29	-0.46	0.8	-0.17	-0.28
2	-0.16	4.27	2.52	0.65	2.45	1.65	2.56	2.96	3.6	1.4
3	-0.22	4.52	2.52	2.13	1.42	1.49	0.74	2.96	0.92	1.4
4	-0.34	0.34	6.71	3.34	.	1.31	2.22	1.12	0.61	0.92
5	1.48	4.52	2.52	0.65	1.42	0.91	.	.	0.24	0.36
6	1.45	0.34	6.08	1.87	2.78	1.26	2.33	1.33	0.71	1.08
7	1.17	0.64	2.52	0.65	-0.57	1.77	2.67	2.13	1.02	5.52
8	1.08	.	.	2.13	.	0.74	.	0.28	0.09	0.12
9	0.95	.	.	2.13	.	1.14	.	1	1.13	0.68
10	0.67	5.05	2.52	2.13	1.42	1.43	0.66	3.66	1.39	1.08
11	0	4.27	2.52	2.13	2.78	1.43	0.44	2.64	0.71	1.08
12	-0.12	4.81	2.52	2.13	1.42	2.95	1.42	2.02	0.56	-0.2
13	0	-1.56	.	.	.	.	1.82	2.14	0.24	.
14	-1	0.39	.	.	.	.	0.79	0.3	1.64	-1.08
15	-1.16	2.36	.	.	.	.	1.59	2.32	0.04	.
16	-1.46	2.75	.	.	.	.	-0.53	-2.06	-0.79	-2.91
17	-1.6	4.27	.	.	.	.	-0.53	-2.06	-1.88	-4.1
18	-1.7	-1.57	.	.	.	.	0.39	-0.42	-0.13	-1.64
19	-0.13	1.87	.	.	.	.	-0.43	1.12	-0.27	.
20	-0.59	4.15	5	0.65	2.45	-0.12	-0.89	0.3	-0.69	-1.08
21	-0.68	4.27	.	.	.	.	1.24	1.12	-0.27	-0.44
22	-0.68	2.36	.	.	.	.	0.79	0.3	-0.69	.
23	-0.7	-2.01	.	.	.	.	3.33	.	.	.
24	-0.82	2.45	3.47	0.7	2.45	-0.12	-0.89	-0.13	-0.69	-1.08
25	-0.89	0.77	.	.	.	.	1.82	2.14	1.17	0.36
26	-0.89	-0.04	.	.	.	.	0.39	-0.42	-1.05	.
	Correl	**0.078**	**0.201**	**0.040**	**0.405**	**0.029**	**0.265**	**0.459**	**0.339**	**0.638**

The model no 10 demonstrated the good predictictive ability for against test and hence was chosen as the best model that explains the estrogen receptor binding affinity activity.

CONCLUSION

The QSAR studies have been successfully applied to a set of compounds to generate essential structural and physicochemical requirements in term of common biophoric sities (Pharmacophore) and secondary sites for binding and interacting with H_1 receptors. The Apex 3D model reveals regions in 3D space around these ligand and provides a hypothetical picture of the main chemical features viz. Pi_population, charge_Het and Don_01 for first biophoric site (BS1) and Cycle_size and Pi_electron for second and third biophoric sites (BS2 & BS3) respectively. The analysis also shows the significant correlation of hydrophobic and steric factors with biological activity.

Catalyst has led to generation of bioactive conformations and hypothesis obtained identified important features (such as hydrophobic, aromatic hydrophobic, hydrogen bond acceptor and steric refractivity) of the surface accessible models. The hydrohobic and positive ionizable features are the minimum components of an effective estrogen antagonistic binding hypothesis.

The model has led to predictions, indentification of the pharmacophore and improvement of understanding of receptor topography in terms of interaction or binding sites.

REFERENCES

1. Topliss, J.G., ***J. Med. Chem.,*** **15**,1006 (1972).

2. Ariens, E.J., in E.J. Ariens (Ed.), ***Drug Design,*** Academic press, New York, ***Vol 1***, pp.1 (1971).

3. Karolkovas, A., ***Essentials of Molecular Pharmacology***, wiley, New York (1970).

4. Crum-Brown, A and Fraser, T.R., ***Trans., Roy. Soc., 25,*** 693 (1868-869).

5. Hansch, C. and Leo, A., ***Substituent constants for correlation analysis in chemistry and Biology,*** Wiley-Interscience, New York (1979).

6. Gupta, S.P., ***QSAR studies on drugs acting at the central nervous system, Chem. Rev., 89***, 1765(1989).

7. Taft, R.W., ***Steric effect in organic chemistry*** edited by Newman, M.S., Wiley, New York, pp. 556 (1956).

8. Fujita, T., Imasa, J. and Hansch, C., ***J. Med. Chem., 8,*** 150, (1965).

10. Hansch, C., Leo, A., Unger, S. H., Kim, K.H., Nikaitani, D. and Lien, E. J., ***J. Med. Chem., 16***, 1207 (1973).

11. Kuntz, I.D., ***Structure based strategies for drug design and discovery science, 257***, 1078 -1082 (1992).

12. Ariens, E.J. in E.J. Ariens(Ed.), ***Drug Design,*** Academic Press, New York, **Vol.1**, pp.1 (1971).

13. Martin, Y.C., **Advanced in Methodology of Quantitative Drug Design, Drug Design, Vol. III**, (Ed.) Ariens, E.J., Academic Press, New York, pp. 2-68 (1979).

14. Boyd. D.B., *Success of computer assisted molecular design, J.Comput. Chem.*, 1, 355-371 (1990).

15. Navia, M.A. and Murcko, M.A., ***Curr. Opin. Struct. Bid.,*** 202-210, (1992).

15a. Bugg, C.E., Carlson, W.M. and Montgemery, J.A., **Science, Ann., 92,** 98 (1993).

15b. Peters, R. and Mekinsky, R.C., **Biotechnology,12**, 147-150 (1994).

16. Free, S. M. and Wilson, J. W., ***J. Med. Chem., 7***, 395, (1964).

17. Bocek, K., Kopecky, J., Krivucova, M. and Vlachova, D., ***Expermentia***, **20**, 667 (1964).

18. Taft, R.W., ***Steric effect in organic chemistry*** edited by Newman, M.S. Wiley New York (1972).

19. Fukui, K., Yonezawa, T. and Nagata, C., **Bull. J. Chem. Phys., 20**, 722 (1952).

21. Dunn, W. J., ***J., Med. Chem.***, **12**, 109 (1977).

22. Moriguchi, I, Kanada, Y and Komatsu, K., ***Chem. Pharma. Bull. (Japan), 24,*** 1799, (1976).

23. Martin Y. C., ***Advanced in methodology of quantitative drug design, Vol. VIII***, (Ed.), Ariens, E. J., Academic Press, New York, pp. 2-8 (1979).

24. Mager, H. and Mager, P., ***Validation of QSARs, Some Reflections***, ***Quan. Struct.-Act. Relat. 2***, 518-21 (1992).

25. Draper, N. and Smith H., ***Applied Regression Analysis,*** John Wiley and sons, New York, 1(1981).

26. Mitler, A. J.,***Subset Selection in Regression***, Chapman and Hall, London (1990)

27. Bolsley, D. A., **Conditioning Diagnostic, Colinearity and weak data in Regression,** John Wiley and Sons, New York, (1991).

28. Mager, H., **Modern Regression Analysis, Mit Anwendungsbei Spielen aus Biologic, Chemic and Medizin Salle and Sauerlander**, Frankfurt, A. M. (1981).

29. Weld, S., ***Validation of QSARs Quan. Struct. Act. Relat., 10,*** 191 (1991)

30. Wold, S. and Dunn III, W. J., ***Multivariate Quantitative Structure Activity Relationship (QSAR) : conditions for their applicability, J. Chem. Inf. Comput. Sci., 23***, 6, (1983).

31. Staodte, R. G. and Sheather, S. J., ***Robust Estimation and Testing***, Wiley, New York, (1990).

32. Gupta, S. P., ***QSAR Studies on Drug Acting on the Central Nervous System, Chem. Rev., 89,*** 1765 (1989).

33. Mager, H., ***The Application of PHL concept to QSAR: Detection of data deficiencies, cross validation, detection of Alien effects, Quan. Struct. Act. Relat., 3***,147, (1984).

34. Gupta S. P., and Prabhakar , Y.S. , ***J. Sci., Endus. Res., 44,*** 189 (1985).

35. Butterworths, W.G., ***Quantum Phamacology***, 19 (1977).

36. Langley, J. N., ***J. Physiol., London, 1.*** 339-369 (1878).

37. Fischer. E., ***Ber. Dtsch. Chem. Ges., 27***, 2985-93 (1894).

38. Dean P.M., ***Molecular Foundations of Drug Receptor Interaction,*** Cambridge University press, Cambridge (1987).

39. Ariens, E. J. ***Drug Design, Vol I,*** Ariens E. J. ed., Academic Press, New York, pp.1-270 (1971).

40. Emmett. J. C., Ed., ***Membranes and Receptor, Vol. 3*** of comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds. Hansch, C. Sammes, P. G. and Taylor, J.B., cds Pergamon press, Oxford (1990).

41. Williams, R.J.P., ***Angew. Chem.,*** Int. (Ed.) Eng., ***16,*** 766-77 (1977).

42. Roberts, G.C.K., ***Quantitative Approaches to Drug Design Phamacochem. Libr. 6.***, Elsevier, Amsterdam (1983).

43. Blundell, T. L. and Johnson L. N., ***Protein Crystallography***, Academic press, London (1976).

44. Branden, C., and Tooze. J., ***Introduction to Protein Structure***, Garland Publishing Company, New York (1991).

45. Bode. W., and Hubes. R., ***Eur. J. Biochem., 204,*** 433-51 (1992).

46. Herbette. L. G. Chester. D.W., and Rhodes, D.G.,***J. Biophys., 49,*** 91-94 (1986).

47. Mason, R.P., Rhades. D.G. and Herbette, L.G., ***J. Med. Chem., 34***, 869-77 (1991).

48. Franke, R., ***QSAR and Strategies in the Design of Bioactive Compounds,*** edited by J.K.Seydel, pp.59(1960).

49. Gumbown. A. and Fraser.T.R., ***Trans. Roy. Soc, Edinburgh, 25,*** 151-203, 693-739 (1868-69).

50. Richet, M.C., *Compt. Rend. Soc. Biol., Paris, 45,* 775-76 (1893)

51. Meyer, H., **Arch. Exp. Path. Pharm., 42,** 109-118 (1899).

52. Overton E., ***Studien uber die Narkose, Zugleich ein Beitrag Zur Allegmeiner Pharmakologic,*** G. Fishcher, Jena, 1901, English translation by Lipnick, R. L., ed., studies on Narcosis. Charles Ernest Overton, Chapman and Hall, London (1991).

53. Ferguson, J., ***Proc. Roy. Soc., Ser., 127,*** 387 - 404 (1939).

54. Bruice, T.C., Kharasch, N. and Winzer, R.J., ***Arch. Biochem. Biophys., 62,***. 305 -317 (1956).

55. Hansch. C. and Fujita. T., ***J. Am. Chem. Soc., 86,*** 1616-1626 (1964).

56. Free, S. M. and Wilson, J. W., ***J. Med. Chem., 7,*** 395-399 (1964).

57. Fujita, T., Iniasa, I. and Hansch, C., ***J. Med. Chem., 8,*** 150 (1965).

58. Purcell, W. P. and Clayton, J. M., ***J. Med. Chem., 17,*** 981 (1974).

59. Ban, T. and Fujita. T., ***J. Med. Chem., 12,*** B53 (1969).

60. Hansch., C., Maloney. P. P., Fujita. T., and Muir, R.M., ***Nature, 194,*** 178-180 (1962).

61. Hansch, C., ***Acta. Chem. Res., 2,*** 232-39 (1969).

62. Hansch,.C., ***Drug design, Vol I,*** Ariens. E.J. (Ed.), Academic Press, New York, pp. 271-342 (1971).

63. Topliss, J.G., ***Quantitative Structure Activity Relationship of Drugs,*** Academic Press, London (1984).

64. Fujita, T. and Ban, T., ***J. Med. Chem., 14,*** 148-152 (1971).

65. Kubinyi, H., ***Quantitative Drug Design, Vol. 4*** of Comprehensive Medicinal Chemistry, The rational Design Mechanistic study and Therapeutic application of chemical compound, eds., Pergaman Press, Oxford, 589-643 (1990).

66. Kubinyi, H., ***Quant. Struct. Act. Relat., 7,*** 121-133 (1988).

67. Schaad. I.J., Hess, B. A., Purcell, W. P., Cammarata, A., Frank, R. and Kubinyi, H., ***J. Med. Chem., 24,*** 900-901 (1981).

68. Cramer, R.D., in M. Tichy (Ed.), ***Quantitative Structure Activity Relationships***, Akademiai Kiado, Budapest (1969).

69. Cramer, R. D., Redl, G. and Berkoff, C.E. , ***Chem. Soc. Rev., 3,*** 273 (1974).

70. Redl. G., Crammer, R. D. and Berkoff, C.E., ***J. Med. Chem., 17***, 533 (1974).

71. Weiner, P. H. , Malinowski, E. R. and Levinstone, A. R., ***J. Phys. Chem., 74,*** 4537, (1970)

72. Weiner, P.H. and Howery, D.G., ***Anal. Chem., 44,*** 1189 (1972).

73. Weiner, P. H., ***J. Amer. Chem. Soc., 95***, 5845, (1973).

74. Weiner, M. L. and Weiner, P. H., **J. Med. Chem., 16**, 655 (1972).

75. Malinkowski, E. R., Anal. Chem., 49, 606 (1977).

76. Malinkowski, E.R.and Howery, D.G., ***Factor Analysis in Chemistry,*** Wiley, New York (1980).

77. Franke, R., Dove, S. and Kuhne, R., ***Eur. J. Med. Chem., 14,*** 363 (1979).

78. Malinowski, E. R., ***Anal. Chem., 49,*** 612 (1977).

79. Skelnar, H. and Jager, J., ***Int. J. Quant. chem., 16***, 467 (1979).

80. Skelnar, H., Jager, J. and Susmetch, R., ***Quantitatively Structure Activity Analysis*** (Ed.) R. Frank and P. Oehme, Academic-verlag, Berlin, pp. 303 (1978).

81. Schonfelder, D and Franke, R., In R. Frank and P. Oehme, ***Quantitatively Structure Activity Analysis*** (Ed.), Academic-verlag, Berlin, pp. 303 (1978).

82. Krause, G., Franke, R. and Vasciliev, G. N., ***Physiol, Plant, 174,*** 128 (1979).

83. Prakash, G. and Hoblenett, E. H., ***J. Med. Chem., 21,*** 369 (1978).

84. Dunn. W.J.and Greenberg. M.J., **J. Pharm. Sci., 66**, 1416 (1977).

85. Henry. D.R. and Block, J. H., **J. Med. Chem., 22**, 465 (1979).

86. Krause, G., Franke, R and Vassiliev, G.N., ***Physiol Plant, 174,*** 128 (1979).

87. Moriguchi, I., Komatsu, K and Matsushi, Y., ***J. Med. Chem., 23***, 20 (1980).

88. Good A. C., Stephan, J. P. and Richards, W.G., QSAR's from imilarity matrices, technique validation and application in the comparission of different similarity. Evaluation methods , ***J. Med. Chem., 36,*** 2929-37 (1993).

89. Klopman, G., ***J. Am. Chem. Soc.,106,*** 7315-7321 (1984).

90. Gombar,V.K., Borgstedt, H.H., Enslein, K., Hart, J.B., Blake, B.W., ***Quant. Struct. Act. Relat.,*** 10, 306-332 (1991).

91. Klopman, G., ***Environ. Health Perspect., 61,*** 269-274 (1985).

92. Klopman, G., Dimayuga, M.L., ***Comput. Aided Mol. Design 4,*** 117-130 (1990).

93. Thornber, C.W. , ***Chem. Soc. Rev., 8,*** 563-580 (1979).

94. H. Kubinyi, (Ed.), ***3D - QSAR in Drug design: Theory, Methods and Applications,*** Escom Science, Lieden (1993).

95. Golender, V.E. , Vorpagel, E.R. , ***Computer-Assisted Pharmacophore Identification, 3D QSAR in Drug Design,*** (H.Kubinyi, Ed.), ESCOM, Leiden pp. 137-149 (1993).

96. Cramer, R.D., Patterson, D.E. and Bunce, J.D., ***Comparative molecular field analysis (COMFA),*** 1, Effect of shape on binding of steroids to carrier proteins, ***J. Am. Chem. Soc., 110*** , 5969 (1988).

97. Thibuat, U, ***Application of COMFA and related 3D QSAR approaches***, ***3D-QSAR in Drug Design : Theory , Methods and Applications*** (H. Kubinyi, Ed.) ESCOM Science, Lieden, pp.66 (1993).

98. Hlopfinger, A. J., ***A QSAR investigation of dihydrofolate, reductose inhibition by baker triazines based upon molecular shape analysis, J. Am. Chem. Soc., 120***, 7196 (1980).

99. Ungar, S.H., ***QSAR in design of Bioactive Compounds*** (K.Kuchar. Ed.) J. R. Prous. Barcelona, pp.1 (1984).

100. Humblet, C. , Marshall. G.R., ***Drug. Dev.Res., 1,*** 409-434 (1981).

101. Doweyko, A. M., ***The hypothetical active sites from data on inhibitor molecules, J. Med. Chem, 31,*** 1396 (1988).

102. Catalyst, BIOSYM/ MSI, Inc. San Diego CA (1995).

103. Apex-3D, BIOSYM Technologies, Inc. San Diego, CA, VSA (1993).

104. Manhold, R., Krosgsgaand-Larsen,R., Timmerman, ***QSAR Analysis,*** VCH, Weinheim, New York (1972).

105. Hansch,C., Unger,S. H. Forsythe,A.B., ***J. Med.Chem., 16,*** 1217-1222 (1973).

106. Dill , K. A., ***Science, 250***, 297 (1990).

107. Taylor, P. J., ***The Rational Design, Mechanistic Study and Therapeutic Application of Chemical Compounds***, eds., Pergaman Press, Oxford, pp. 241-294 (1990).

108. Rekker, R.F., ***Pharenacochem. Libr.1***, Elsevier, Amsterdam (1977).

109. Hansch,C., Leo,A., ***Substitution Constants for Correlation Analysis in Chemistry and Biology,*** Wiley, New York (1970).

110. Taft Jr, R., W., ***Steric Effects in Organic Chemistry,*** Newmann, M.S. Ed., Wiley (1980).

111. Hammet, L. P., ***Physical Organic Chemistry, Reaction Rates, Equilibiria and Mechanism***, 2nd Ed. Mcgraw -Hill, New York (1970).

112. Perrin, D.D., Dempsey B., and Serjeant. E.P., ***pka prediction for organic acids and bases***, Chapmann and Hall, London (1981).

113. Chapman, N. B., and Shorter, J. (Ed.) ***Advances in Linear Free Energy Relationship,*** Plenum Press, London (1972).

114. Dunn III, W.J., ***Eur.J.Med. Chem., 12,*** 109-112 (1977).

115. Blaney, J.M., Hansch, C., **Quantitative Drug Design, Vol. IV** of Comprehensive Medicinal Chemistry, pp. 459-496 (1990).

116. Leo. A., Hansch, C. and Elkins, D., ***Chem. Rev., 71***, 525-616 (1971).

117. Moriguchi, I., Hirono, S., Liu, Q., and Nakagome , I., ***Chem. Pharm. Bull., Japan, 40***, 127-130 (1996).

118. Van. de waterbeemed H. and Testa, B., ***Adv. Drug Res.,16,*** 85-225 (1987).

119. Vishwanadhan, V. V. , Ghose, A.K. and Revenkar, G.R. , ***J. Chem. Inf. Comput. Sci., 29,*** 163-172 (1979).

120. Verloop.A., ***Drug Design,*** J. Motoc, Ed., Springer, Verlag, Berlin (1983).

121. Bonde, A., ***J. Phys. Chem., 69,*** 3284 (1965).

123. Hinze, J., Whitehead, M.A., Jaffe, H. H., ***J. Am. Chem. Sec., 85,*** 148 (1963).

124. Sanderson, R. T., ***Chemical Periodicity*** , Reinholds , New York (1960).

125. Draper, N.R. , Smith, H., ***Applied Regression Analysis,*** Wiley, New York (1966).

126. Daniel, C, Wood, F.S., ***Fitting Equations to Data,*** 2nd Ed., New York (1980).

127. Sneduuor, G.W., Cohran, W.C., ***Statistical Method,*** 6th Ed., The Iowa State University Press, Amesterdam (1967).

128. Martin, Y.C. , Holland, J.B. , Jarboc, C.H. and Plotnikoff, N., ***J. Med. Chem., 17,*** 409-413 (1974).

129. Smith, C.C., Genther, C.S. and Coats, E.A., ***Eur.J. Med. Chem., 14,*** 271-276 (1979).

130. Lewis, D.F.V., Loannides, C. and Parke, D.V., ***Mutagenesis, 5***, 433-435 (1990)

131. Moriguchi, I., Komatsu, K. and Masushita, Y., **Anal. Chem. Acta 133**, 625-636 (1981).

132. Moriguchi, I., Komatsu, K. and Matsushita, Y., **Anal. Chem. Acta., 133**, 625-636 (1981)

133. Tokahashi, Y., Miyashita, Y. and Sasaki, S., ***Bunseki., Kagaku, 33,*** 487-494 (1984).

134. Dunn III, W.J. and Wold, S., ***J. Med. Chem., 21,*** 1001-1007 (1978).

135. Norden, B., Edlund, U., Johnels, D. and Weld, S., ***Quant. Struct. Act. Relat.***, *2*, 73-76 (1983).

136. McFarland, J.W. and Gans, D.J., ***Drug Information, 24,*** 705-711 (1990).

137. Hansch, C., ***Acta. Chem. Res., 26,*** 147-153 (1993).

138. Katzenellenbogen , J. A. and Katzenellenbogen, B. S., ***Chem. Biol., 3,*** .529 (1996).

139. Hormutz, K. B. , Jackson, T. A., Rain D.L. , Richer, J. K., Takimato, G. S. and Tung, L., ***Mol. Endocrinol., 10***, 1167 (1996).

140. Mosselman, S., Polman. J. and Dijkema, R., ***Litt., 49,*** 392 (1996)

141. Katzenellenbogen, B.S. and Korach, K.S., ***Endocrinology, 138,*** 861(1997).

142. Kuiper, G. G. J. M., Carlson, B, Grandien, K., Enmark, E., Haggblad, J., Nilsson, S. and Gustafsson, I., ***Endocrinology, 138,*** 863 (1997).

143. Couse, J. F., Lindzey, J., Grandein, K., Gustafsson, J. A. and Korach, K. S., *Endocrinology, 138,* 4613 (1997).

144. Kuiper, G. G. J. M. and Gustafsson, J. A., ***FEBS Lett., 87,*** 410 (1997).

145. Katzennellenbogen, J. A., O' malley, B.W. and KatznellenBogen, B. S., *Mol. Endocrinol., 10,*119 (1996).

146. Lobo R. A., ***In estrogens and anti-esterogens, basic and clinical aspects***, Lindsay, R., Dempester, D. W. and Jordon, V. C., Eds., Lippincott-Raven: Philadelphia, pp. 63 (1997),

147. Sheruin, B. B. ***In estrogens and antiestrogens,*** Lindsay, R., Dempster, D. W. and Jordon, V. C., Eds., Basic and Clinical aspects, Lippincott-Ravan: Philadelphia, pp.75-87 (1997).

148. Compston, J. E. , ***Lancet., 350***, 676 (1997),

149. Stanford, J. L., Weiss, N. S., Voigt, L. F., Daling, J. R. , Habel, L. A. and Rosing, M. A. , ***J. Am. Med. Assoc., 274,*** 137 1995.

150. Lin, X. and Huebner, V., ***Curr. Opin. Drug Discovery Dev.,3***,383 (2000).

151. Grese, T. A. and Dodge, J. A, ***Curr. Pharm. Des.***, **4,** 71 (1998).

152. Tonetti, D. A. and Jordon, V. C., Targeted anti-estrogens to treat and prevent diseases in women., ***Mol. Med. Today, 2***, 218-223 (1996).

153. Jordon V. C., A current view of tamonifen for the treatment and prevention of breast cancer., ***Br. J. Pharmacol., 110,*** 507-517(1993).

154. Grese, T. A. and Dodge, J. A., Cuss., ***Pharm. Des., 4***,71. (1998).

155. Kuiper, G.G.J.M., Carlsson, B., Grandien, K., Enmerk, E., Haggblad, J., Nilsson, S. and Gustafsson, J., **Endocrinology, 138,** 863 (1997).

156. Hochberg, R.B., ***Science, 205,*** 1138 (1979),

157. Gosulami, R., Harry, S.G., Heiman, D.F., Katzenellenbogen, J.A. J. Med. Chem. 1980, 23 (1994)

158. Todeschini, R and Consonni, V., ***Various Molecular Descriptors,*** A Handbook of Molecular Descriptors wiley-VCH, weinheim (Germany). (2000)

159. Carlson, K.E., Chai, I., Gee, A., Katzenbogen, B.S. and Katzeneller bogen, J.A., ***J. Med. Chem, , 42,*** 2456-68 (1999).

160. Meyers, M.J., Sun.J., Carlson. K.E., Katzenellenbogen, B.S. and Katzenellenbogen, J.A., ***J. Med. Chem., 42,*** 2456-68 (1999).

161. Stauffer, S.R. , Tedesco, R., Nishiguchi, G., Carlson, K., Sun, J. and Katzenellenbagen , J.A. , ***J. Med. Chem. 43,*** 4934-47 (2000),.

162. Labaree, D.C. , Reynolds, T.Y. and Hochberg, R.B., ***J. Med. Chem.,*** , **44,** 1802-1814 (2001).

163. Sarrel, P.M., ***Obstet. Gynecol., 318***, 96-100 (1990),.

164. Nishiguchi, G.A., Rodrigruez, A.L. and Katzennellenbogen, J.A., ***Bio. Med. Chem. Lett. , 12,*** 947-950.(2002)

165. Katzenellenbogen, J.A., ***Breast Cancer Res., 2***, 335 (2000)

166. Carlson, K.E. , Choi, I., Gev A. , Katzenellenbogen, B.S. and Katzellenbogen, J.A. , ***Biochemistry., 36***, 148-897 (1997)

167. Wagner, R.L., Apriletti, J.W., Mc. Grath, M. E., West, B. L., Baxter, J.D. and Fletterick, R. J., ***Nature, 378,*** 690(1995)

168. Gantcher, T.G., Ali, H. and VanLier, **J. Med. Chem., 37,** 4164(1994).

169. Poirot, M., Medina, P.D. , Delarue. F. Perie, J.J., Klaebe, A. and Faye, J.C. , *Bioorg. Med. Chem., 8*, 2007-2016 (2000).

170. Faye, J.C, Jozan, S., Redeilh, G., Baulieu, E.E. and Bayaqqrd F., *Proc. Natl. Acad. Sci., USA, 80*,3158 (1983).

171. Watts, C.K. Murphy, L.C. and Sutherland, R.L., *J. Biol. Chem., 259,* 4223. (1984).

172. Brandes, L.J., *Biochem. Biophys. Res. Commun., 124*, 244 (1984).

173. Delarue, F., Kedjouor, B., Mesange, F., Bayard, F., Faye J.C., poirot, M. *Biochem Pharmacol, 58*, 1927(1999).

174. Kedjouor, B., Daunes, S., Vilner, B.J. Bomen, W.D., Klaebe, A., Faye, J.C. and Poirot, M., *Biochem. Pharmacol., 58,* 1927. (1999).

175. Messange, F., Delarue, F., Puel, J., Bayard, F. and Faye, J.C., *Mol. Pharmacol., 50,* 75 (1996).

176. Henke, B.R., Drewry, D.H. , Jones, S.A., Steuart, E.L., Weaver, S.L. and Wiethe, R.W., *Bioorg Med. Chem. Lett.*, 11, 1939-1942 (2001).

177. Burry, P.S., Christiansen, L.B. , Jacobsen, P., Jorgensen, A.K., Naerum, L., Bain. S., Fledilius, C., Gissel, B., Hansen, B.S., Koresgaard, N., Thorpe, S.M. and Wassermann, K., *Bioorg. Med. Chem., 10,* 125-145. (2002).

178. Thorpe, S.M., *Breast Cancer Res., Treat., 9*,175 (1987)

179. Andersen, H.R. , Anderson A.M. , Arnold S.F., Autrup, H, Barfoed, M,, Beresford, N.A., Bjerregaard, P., Chistiansen, L.B., Gissel, B., Hummel, R., Liffers, H., Mclachtan, J., Moller, A., Nielsen, J.B. Olea, N, Oles-karasko, A., Pakdel, F., pederson K.L., Perez, P., Seakkeback, N.E.,

Sonnenschein, C. Soto, A.M., Sumpter, J.P., Thorpe, S.M. and Grandjean, P. *Environ., Health Perspect., 107*,89 (1999).

180. Anstead, G. M., Carlson, K.E. and Katzenellenbogen , J.A., ***Steroids,*** *62*, 268 (1997).

181. Tong, W. and Perkins, R., ***Endocrinology, 138***, 4022 (1997).

182. Bignon, E., Pons, M., Crastes de Paulet, A., Dore, J.C., Gilbert, J., Abecassis, J., Miquel, J.F., Ojasso, T. and Raynaud, J.P., ***J. Med. Chem., 32***, 2092 (1989).

183. Gilbert, J., Fuentes, M., Ojasso, T., Dore, J.C. and Pons, M., ***J .Med. Chem., 40***, 1104 (1997).

184. Anstead, G.M., Wilson, S.R. and and Katzenellenbogen, J.A., ***J. Med. Chem., 32***, 2163 (1989).

185. Anstead, G.M. and Katzenellenbogen, J.A., ***J. Med. Chem., 31,*** 1754 (1989).

186. Stewart, J.J.P., ***MOPAC : A semiempirical molecular orbital program, J. Comput. Aided Mol. Des., 4,*** 1 (1990).

187. Ghose, A. K. and Crippen, J. M., ***J. Chem. Inf. Comput. Sci., 27,*** 21 (1987)

188. Golennder, V.E. and Rosenbit, A.B., ***Logical and Combinatorial algorithms drug design,*** Research studies press, Letchworth, U.K. (1999).

189. Pandya, T., Pandey, S. K., Tiwari, M., Chaturvedi, S. C. and Saxena, A. K., ***Bioorg, Med. Chem., 9,*** 291 (2001).

केन्द्रीय औषधि अनुसंधान संस्थान
छत्तर मंजिल, पोस्ट बाक्स सं० 173 लखनऊ - 226 001 (भारत)
Central Drug Research Institute
Chattar Manzil P.O. Box 173, Lucknow - 226 001 (India)
Phone: 223286, 210932, 212411-18(PABX) (Res.) 321625, 322552
Fax: 091-(522)-223405/223938/229504 **Gram:** CENDRUG
Email: drcmg@satyam.net.in, root@cscdri.ren.nic.in, cmg@gateway.cdri.res.in
Web: htpp://www.cdriindia.org

डॉ० सी०एम० गुप्ता एफ एन ए, एफ ए एससी, एफ एन ए एससी, एफ ए एम एस
निदेशक

Dr. C.M. GUPTA FNA, FASc, FNASc, FAMS
Director

May 9, 2002

TO WHOM IT MAY CONCERN

This is to certify that **Ms. Soumya Srivastava,** a student of Allahabad University, Allahabad has undergone training in our Division of Medicinal Chemistry from 2.4.2002 to 1.5.2002. During the period of training she has worked on 2-D QSAR and also done some work on 3-D QSAR.

Ms. Srivastava is a sincere and diligent worker. I wish her every success in her career.

15/5/02

(C.M. Gupta)

केन्द्रीय औषधि अनुसंधान संस्थान
छतर मंजिल पैलेस, पोस्ट बाक्स नं० 173, लखनऊ (भारत)
Central Drug Research Institute
Chattar Manzil , Post Box No 173, Lucknow- 226 001(India)

डा० सी०एम० गुप्ता, एफ०एन०ए०, एफ०ए०एस०सी०, एफ०ए०एस०टी०, एफ०ए०एम०एस०, एफ०टी०डब्लू०ए०एस
Dr. C.M. GUPTA, FNA, FASc, FNASc, FAMS, FTWAS
निदेशक
Director

सी० एस० आई० आर० हीरक जयंती समारोह
26 सितम्बर 2002 से 26 सितम्बर 2003
CSIR Diamond Jubilee Celebrations
Sept. 26, 2002 to Sept. 26, 2003

Nov. 7, 2002

TO WHOM IT MAY CONCERN

This is to certify that **Ms. Soumya Srivastava,** a student of M.Sc., Allahabad University, Allahabad has undergone training in our Division of Medicinal Chemistry w.e.f. 1.10.02 to 31.10.02. During the period of training she has learnt 3D QSAR and molecular modeling.

Ms. Srivastava is sincere and diligent worker. I wish her every success in her career.

(C.M. Gupta)

Phone: 91-522-223286, 210932, 91-522-212411-18 (PABX) 4222 Extn. **Resi:** 91-522-322552, 321625
Fax: 91-522-223405/223938 **Email:** drcmg@satyam.net.in / drcmg@rediffmail.com **Web:** www.cdriindia.org